The Dissertation Committee for Riddhiman Kannan Garge Certifies that this is the approved version of the following Dissertation:

Of Yeast and Men: Insights into evolution and human health from 1 billion years of divergence

Committee:

Edward Marcotte, Supervisor

John B. Wallingford

Claus O. Wilke

Arlen W. Johnson

Jeffrey E. Barrick

Of Yeast and Men: Insights into evolution and human health from 1 billion years of divergence

by

Riddhiman Kannan Garge

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2021

Dedication

To my grandparents, *Thatha* and *Patti*, for always nurturing my curiosity.

Acknowledgements

This dissertation would not have been possible without the help of mentors, colleagues, friends, and family. First and foremost, I owe huge thanks to my ever inspiring and brilliant advisor Edward Marcotte, for letting me run wild and free with projects in his fantastic laboratory and (more importantly) guiding me when I needed it the most. I could not have imagined a more enjoyable graduate school experience.

I will always be grateful to my committee members for their mentorship and valuable feedback which often generated interesting ideas and fruitful collaborations over the years. Particularly, I'd like to thank John Wallingford for always pushing me to communicate my work with broad scientific contexts and Claus Wilke for enforcing the fundamentals of data visualization early. I must thank Andy Ellington for pushing me to think big and beyond the bounds of my projects. While Andy's mentoring style initially seemed counterintuitive, it has often motivated me to pursue exciting projects. An understated perk of being in Edward's lab is the opportunity to collaborate with several groups. I have been fortunate to work with 3 wonderful scientists, Maitreya Dunham, Michael Polymenis, and Dannie Durand during their respective sabbaticals in Edward's lab, who all taught me to see science through their eyes and have supported me since.

A laboratory is only as good as the individuals in it and I have been very fortunate to spend my time in Edward's group with some of the most talented and energetic set of researchers to all of whom I am truly grateful. I must start by thanking Aashiq Kachroo and Jon Laurent for willingly taking me under their wing and showing me the widespread utility of Baker's yeast in evolution, systems, and synthetic biology. My graduate work would not have been possible without their constant assistance, encouragement, and help.

I owe huge thanks to the many members of the Marcotte lab that I have crossed paths with. Daniel Boutz for his scientific rigor and advice throughout. I am a better scientist because of him. Andrew Horton for patiently answering all my naïve questions. Fan Tu for proactively pushing me away from stagnation and emphasizing importance of “Inner Peace”. Jimmy Gollihar for his vast wealth of synthetic biology knowledge and his many wagers to prove him wrong. My failure to win any has allowed me to learn so much along the way. Kevin Drew for his advice and allowing me into several of his projects. Benjamin Liebeskind for showing me the power of writing well and addicting me to vinyl record collection. The ever-entertaining Alexander Boulgakov for his constant availability. I would not have joined Edward’s lab if it wasn’t for our crazy and exciting conversations. Angela Bardo for showing me the “correct” ways around a microscope. Anna Battenhouse for her eternal patience and availability. Sophie Curie for being a quick learner meticulously helping me with my projects. Chip McElroy and his Live Oak team for showing me that interesting scientific questions are not restricted to the lab and always keeping me supplied with “samples”. Finally, Christopher Yellman for sharing his knowledge on “the awesome power of yeast genetics”. Those who are not named here are by no means forgotten but will always be appreciated.

I owe a huge deal of gratitude to my undergraduate mentors Ramaswamy Subramanian, Jeff Abramson, Vinod Nayak, and Bala Subramanian for letting an engineer with no research background into their respective laboratories and being strong guides ever since.

I’d like to thank all my fellow UT friends whose company has made this whole experience wholesome and memorable, especially Elizabeth Gardner, Austin Cole, Seung Woo Ryu, Rohit Satija, Drew Vanderwood, Sean Provost, Jaggannath Swaminathan, Eric Verbeke, Shaunak Kar, Bikash Shrestha, Jojo Mensah, Mayowa Oyedere, Williams

Ozowe, and Ankur Bhambhotta. Outside the lab, I owe a great deal of gratitude to Thejas Kesari, Sriram Sudarsanam, Harsha Shetty, Arjun Shounak, Nayantara Kurpad, Carishma Shetty, Varad Kelkar, and Piyush Jajpuria. Their unwavering friendship has helped keep me sane through it all. I also thank the fellow musicians and bandmates I've worked with over these past few years for helping me achieve some form of work-life balance. Thanks are due to Niranjan Chandrashekaran, Karishma Kaushik, and their respective families for settling me into Austin and being homes away from home.

The biggest of thanks must go to my family for tolerating my idiosyncrasies and being the pillars of support that I often take for granted. My mother, Jaimala Kannan for her infinite excitement and affection. My father, Kannan Subramanian for all he's done in grooming me into who I am today. My brother, Prathyush Kannan for standing by me when the going was tough. Srinivas Ramachandran and Sujatha Jagganathan for being ever-inspiring scientists and oceans of advice. My Ph.D. would not exist if not for my grandparents. I believe this is as much their Ph.D. as it is mine. I will be eternally indebted to my grandfather for his care and affection in nurturing my passion for the Sciences. Though you are not here today to witness the results of science-powered grandparenting, you are always in my thoughts.

Last but by no means least, I'd like to thank Olga Kotlova for her unconditional care, patience, presence, and support even in these rather crazy pandemic-stricken times.

Abstract

Of Yeast and Men: Insights into evolution and human health from 1 billion years of divergence

Riddhiman Kannan Garge, Ph.D.

The University of Texas at Austin, 2021

Supervisor: Edward M. Marcotte

Life on the planet is incredibly diverse and it is often easy to compare and contrast the many features that distinguish any two pairs of species from each other. Despite this diversity, all organisms on Earth share a common origin. This shared ancestry establishes conservation at the core of biology. The concept of conservation (or what's equivalent) across species organizes biology and stems from the natural selection of favorable traits in organisms. Evolutionary conservation extends even to the genetic and molecular level with genes, proteins, and the networks they constitute also sharing common ancestry. This property enables biologists to study conserved genes (orthologs) in simpler model organisms and relate them to their corresponding human equivalents. Despite this, it is largely unclear the extents to which orthologs between species are functionally compatible. The dissertation aims to directly address this question via cross-species gene swaps. By systematically humanizing yeast genes, this dissertation provides insights into how orthologous genes between species functionally diverge and evolve over vast timescales. In chapter one, I present conservation as a powerful organizing principle in biology and the roles orthologous biological systems play in connecting

genotype and phenotype. In chapters two and three, I describe efforts to apply humanized yeast as a platform to study functional divergence in orthologs constituting expanded gene families and examine the trends that underlie them. In chapter four, I describe the synthesis of observations from multiple research threads including humanized yeast, model organisms, evolutionary conservation of biological systems, and global signatures of pesticide resistance to uncover a novel class of antifungals all capable of functioning as vascular disrupting agents. Finally, in chapter five, I discuss the future of cross-species gene swaps, humanized yeast, and their utility to human health and disease.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

"Nothing in Biology Makes Sense Except in the Light of Evolution."

-Theodosius Dobzhansky, 1973

1.1. Evolutionary conservation: An organizing principle in biology

Life on earth is incredibly diverse with the number of species estimated to range between 2 million to 1 trillion (Locey and Lennon 2016). Many species are similar, but no two species are the same and it is often easy to observe and point out all the features that distinguish them from one another (e.g., lifespan, appearance, habitat, diet). Despite this vast diversity, the ancestry of life on earth is shared and deep-rooted in evolution. To make sense of all this diversity one might want to study all the characteristic biological features that distinguish a particular species from its neighbors. Conversely, one might ask the opposite of what is equivalent (conserved) given this vast diversity? This is where the concept of conservation becomes central to biology (and to this thesis). Broadly, a few examples include the process of copying and propagating genetic material from one generation to the next (DNA replication), the ability to convert food to energy (metabolism), and the ability to build macromolecular machines and processes (protein complexes and signaling pathways) involved in various cellular roles. All of these aforementioned examples are central to life and have to be carried out by all organisms in some equivalent form.

This concept of conservation is deeply rooted in all of biology. It organizes biology and arises from the retention of favorable traits in organisms via natural selection. This shared ancestry extends even to the genetic and molecular level with even genes, proteins, pathways, complexes, and their respective interaction networks also sharing a common origin. The concept of conservation is what allows researchers to study cellular and molecular functions of systems in a particular organism and translate it to others. Therefore, studying conserved systems in simpler organisms serve as valuable proxies in supplying key insights into their corresponding human equivalents. This principle often forms the foundation to most contemporary biomedical research including the genetics of human disease, pharmaceutical therapeutics, and drug discovery. The principles of evolutionary conservation and orthology especially in the light of human health and disease inspire this thesis.

1.2. ORTHOLOGOUS SYSTEMS CONNECT GENOTYPE TO PHENOTYPE

One of the ultimate goals in biology is to connect genotype to phenotype. Directly, this connection is probed by observing the consequences (or phenotype) of disrupting the genotype either by introducing mutations or repressing the expression of a gene. Before the genomics era and much before sequences of genes and genomes were available, this was the go-to strategy to infer gene function by association to its phenotype. Since then, the advent of high-throughput DNA sequencing has led to the birth of comparative genomics allowing us to directly identify orthologous genes in distant organisms (including plants, animals, and even microorganisms) and further relate

the effects of their disruption back to humans. These efforts have been largely advanced by the sequencing of the human genome along with forward and reverse genetics. Forward genetics studies allowed scientists to identify the genetic basis behind deleterious phenotypes thereby identifying the corresponding orthologous human genes potentially implicated in disease. On the other hand, reverse genetics enabled scientists to selectively make mutations and/or disruptions in orthologs of known human disease genes in better understood organisms to understand the related phenotypes. Both these approaches gave rise to new model organisms including bacteria (*Escherichia coli*), yeast (*Saccharomyces cerevisiae*), worms (*Caenorhabditis elegans*), fruit flies (*Drosophila melanogaster*), frogs (*Xenopus laevis*) and mice (*Mus musculus*) to study human disease. Studies in these organisms often served as powerful platforms to discover genes implicated in many human diseases. Over the years, these studies have aggregated into a vast ocean of genotype-phenotype associations in model organisms (**Fig. 1.6.1**).

This rich well of data serves as a strong basis to discover and identify seemingly non-obvious relationships between phenotypes in model organisms and human diseases. Motivated by this extensive phenotyping of genetic manipulations in model organisms, McGary *et al*. developed a powerful strategy leveraging these gene-phenotype associations to identify orthologous phenotypes or phenologs. This strategy defines orthologous phenotypes between organisms by systematically overlapping sets of orthologous genes associated with each organism-specific phenotype (McGary *et al.* 2010). In other words, two phenotypes (between species) are said to be phenologs if they share a set of orthologous genes at a statistically significant rate higher than by random chance (**Fig. 1.6.2**). Most interestingly, this approach provides valuable insights into how

conserved gene modules often work in organism-specific contexts and identifies quite inconspicuous relationships between humans and distant species. A few notable relationships discovered included a yeast model for angiogenesis, a worm model for breast cancer, and a plant model for Waardenburg syndrome. Obviously, yeast lack blood vessels, worms don't suffer from breast cancer and plants don't experience Waardenburg syndrome, but these phenotypic relationships are linked by the same sets of orthologous genes these organisms. Phenologs allow for the identification of new candidate genes that are likely implicated in disease opening up new routes to understand disease states. Expanding the definition of phenologs to include phenotypic assays beyond mutant phenotypes can be quite powerful in identifying compounds that therapeutic potential. This was well demonstrated subsequently by Cha *et al*. leveraging the yeast model of angiogenesis to discover that the widely used antifungal Thiabendazole (TBZ) behaved as a vascular disrupting agent allowing for repurposing of the compound as an anti-cancer therapeutic (Cha *et al.* 2012). These methods are only beginning to scratch the surface of the phenotypic relationships that can be established through deeply conserved systems, phenologs, and model organisms. With richer phenotyping, comparative and systems-wide approaches, the number of phenologs will only steadily rise.

1.3. CROSS-SPECIES GENE SWAPS AND HUMANIZED YEAST

This link between model organisms and human health is only reinforced by phenologs making these studies extremely relevant not to mention all the curiosity-driven research in these organisms that have also resulted in invaluable discoveries (like cloning,

genome editing, fluorescent protein tags, and proximity protein labeling) central to contemporary biomedical research. As we continue to learn more about phenotypic associations between species and the conserved genetic systems that underlie them, the central question is whether orthologous genes between species function similarly. Generally, molecular roles of orthologous genes are thought to be conserved between species. But the disruption of the same orthologous genes across species can lead to radically different phenotypic outcomes. For example, mutating the *RB1* gene in humans results in cancer of the retina while disrupting its nematode ortholog causes ectopic vulvae. While both these genes may be executing the same molecular roles, the contexts and environments in which they do so are inherently different between the two species leading to characteristic organism-level phenotypes. The key is that these dissimilar phenotypes are connected by the same set of orthologous genes and it is this principle that has linked studies in model organisms back to the roles of human genes in health and disease. Despite this remarkable power, it is still largely unknown how and to what extent orthologous genes evolve, diverge, and retain function over vast timescales.

Cross-species gene swaps offer direct opportunities to test the ability of functional equivalency among orthologous genes. Such complementation assays have been around since the mid-1980s wherein microbial orthologs of human genes were “humanized”. These early studies used these complementation assays to identify human genes capable of rescuing mutant phenotypes in yeast. Notably, Kataoka *et al.* complemented fully human or chimeric yeast-human versions of *RAS* genes in *rasΔ* strains demonstrating the functional homology between the two species (Kataoka *et al.* 1985). In 1987, Lee and

Nurse screened a library of human cDNAs to identify human orthologs capable of rescuing *CDC2*'s function in fission yeast (Lee and Nurse 1987).

Even though humans and yeast last diverged from their last common ancestor nearly a billion years ago, many genes still share common ancestry. The *Saccharomyces cerevisiae* genome contains ~6,000 protein coding genes while the human genome encodes ~20,000. Comparing orthologs across species, humans and yeast share >2,000 orthologous groups comprising of ~2,000 yeast and ~4,000 human genes (Sonnhammer and Östlund 2015). Baker's yeast (*Saccharomyces cerevisiae*) has been instrumental in advancing our knowledge of basic biology and how they translate to humans. There are several reasons behind this including its short replicative cycles, ease of handling, access to a suite of tools for genetic and biochemical manipulation, genetic tractability and scalability. In 30 years since the early humanization experiments, over 700 yeast genes have been swapped for their human orthologs (Heinicke *et al.* 2007; Cherry *et al.* 2012; Kachroo *et al.* 2015; Hamza *et al.* 2015, 2020; Sun *et al.* 2016; Yang *et al.* 2017; Laurent *et al.* 2020; Garge *et al.* 2020). These complementation studies have been motivated by varied goals including uncovering the roles of human orthologs in simpler yeast contexts (Dunham and Fowler 2013), screening of the functional capacity and pathogenicity of human genetic variants (Hamza *et al.* 2015; Sun *et al.* 2016, 2018; Yang *et al.* 2017) and more broadly leveraging systematic gene swaps to uncover functional divergence across vast timescales across orthologous genes and the families that constitute them (Kachroo *et al.* 2015, 2017; Laurent *et al.* 2020; Garge *et al.* 2020).

Even though there were several reports of human genes replacing their yeast counterparts, it was still unclear how these observations scaled to entire biological

systems and sets of orthologous genes. This inspired the early humanization work in the Marcotte lab. By systematically swapping essential yeast genes for human orthologs, functional complementation could be assayed via simple growth rescue assays. On assaying >400 genes, Kachroo *et al.* found that nearly half of tested essential genes were indeed functionally replaceable revealing remarkable conservation of roles across billion-year timescales (**Fig. 1.6.3**). Additionally, replaceability was driven by modularity meaning that genes involved in common pathways or processes tended to be similarly replaceable or not (Kachroo *et al.* 2015). This is but an entry point to study human processes in yeast. Possessing a census of humanizable yeast genes now allows us to functionally characterize these genes and understand the interactions between them in yeast. The modular nature of replaceability now facilitates the construction of entire human systems and processes in a simpler eukaryotic context. Finally, the scalability and tractability offered by yeast allow for massively parallel functional characterization of allelic variation that can be tolerated by human genes. These topics form the basis for subsequent humanization efforts in this thesis.

1.4. INTRODUCTION TO DISSERTATION

The central theme tying this dissertation together is understanding the conservation of core biological systems, particularly how shared genes (orthologs) between species functionally duplicate, diverge, and evolve over vast timescales. Building off previous work (Kachroo *et al.* 2015, 2017; Laurent *et al.* 2016), this dissertation begins with probing the extents to which orthologs (and co-orthologs)

constituting expanded gene families retain ancestral function through the course of evolution (Chapter 2). Chapter 3 focuses on how ancestral roles are distributed in gene families constituting a core eukaryotic module (i.e., cytoskeleton). Finally, Chapter 4 synthesizes the insights gained from orthologous systems, cross-species gene swaps, and model organisms to uncover a novel class of widely used antifungal compounds all capable of disrupting blood vessels and inhibiting angiogenesis. Besides these main projects, I was involved in several collaborative projects which are described in the appendices. These include applying CRISPR-Cas9 mediated genome editing tools in yeast (Appendix A) and omics-based approaches to understanding the dynamics of biological systems across temporal processes (Appendices B and C).

1.5. PUBLISHED PAPERS

This dissertation consists of an aggregation of my graduate work published during my time at The University of Texas at Austin all of which are listed below chronologically. The papers featured in the main chapters of this dissertation are marked with an asterisk while the others are included in the appendices.

- * Laurent, J.M., Garge, R.K., Teufel, A.I., Wilke, C.O., Kachroo, A.H., and Marcotte, E.M. (2020). Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. *PLOS Biology 18*, e3000627.

- * Garge, R.K., Laurent, J.M., Kachroo, A.H., and Marcotte, E.M. (2020). Systematic Humanization of the Yeast Cytoskeleton Discerns Functionally Replaceable from Divergent Human Genes. *Genetics 215*, 1153–1169.
- * Garge, R.K., Cha, H.J., Lee, C., Gollihar, J.D., Kachroo, A.H., Wallingford, J.B., and Marcotte, E.M. (2020). Antifungal benzimidazoles disrupt vasculature by targeting one of nine β-tubulins. *BioRxiv*.
- Akhmetov, A., Laurent, J.M., Gollihar, J., Gardner, E.C., Garge, R.K., Ellington, A.D., Kachroo, A.H., and Marcotte, E.M. (2018). Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. Bio Protoc *8*.
- Blank, H.M., Papoulas, O., Maitra, N., Garge, R., Kennedy, B.K., Schilling, B., Marcotte, E.M., and Polymenis, M. (2020). Abundances of transcripts, proteins, and metabolites in the cell cycle of budding yeast reveal coordinate control of lipid metabolism. *MBoC 31*, 1069–1084.
- Roberson, E.C., Battenhouse, A.M., Garge, R.K., Tran, N.K., Marcotte, E.M., and Wallingford, J.B. (2021). Spatiotemporal transcriptional dynamics of the cycling mouse oviduct. *BioRxiv*.

1.6 FIGURES

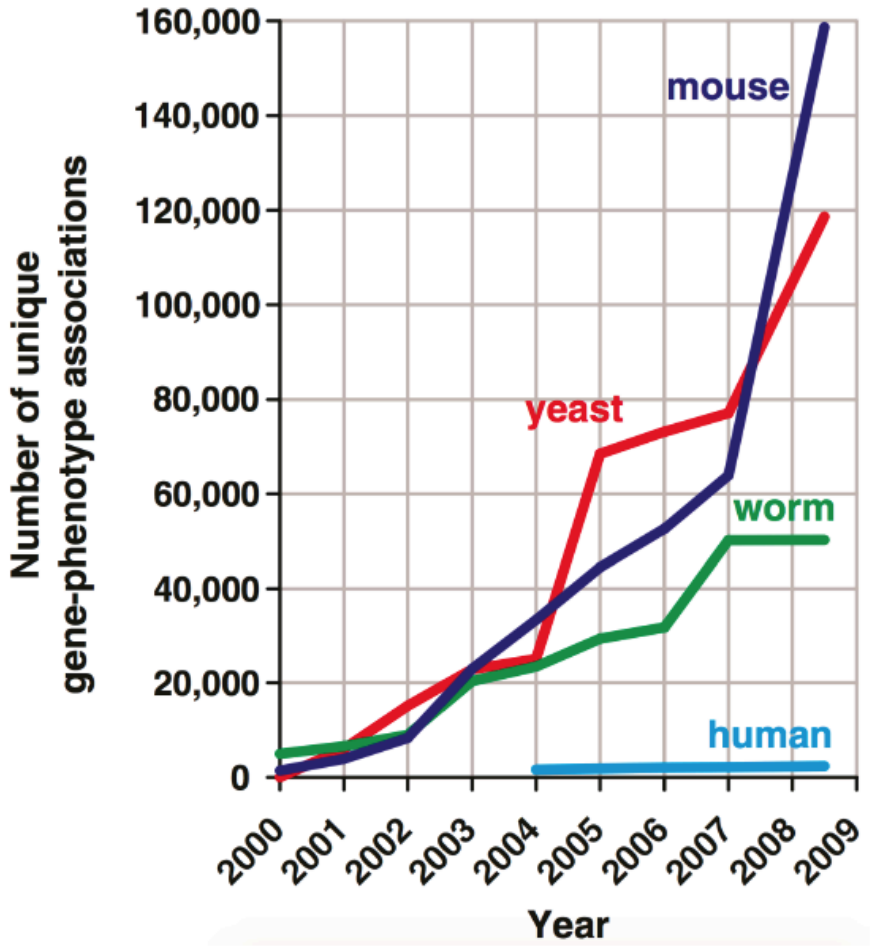


Figure 1.6.1. Rapid growth of genotype-phenotype associations in organisms.

Mapping of model organism genes to human diseases is growing rapidly as compared to human genes to human diseases. Figure adapted from McGary *et al*. (McGary *et al.* 2010).

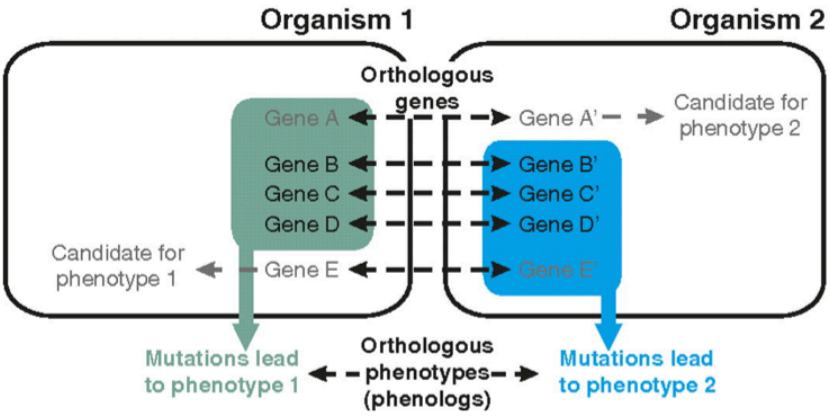


Figure 1.6.2. The concept of phenologs.

Two phenotypes are said to be phenologs when they share a statistically significant set of overlapping orthologous genes. In this figure, the genes in the green box are associated with phenotype 1 in organism 1 while those in blue are associated with phenotype 2 in organism 2. The genes in grey represent candidate genes that can be tested for their respective cross-species phenotypic association. Gene A's ortholog Gene A' as a candidate for phenotype 2 in organism 2 and the ortholog of Gene E' in organism 1 for phenotype 1. Figure adapted from McGary *et al*. (McGary *et al.* 2010).

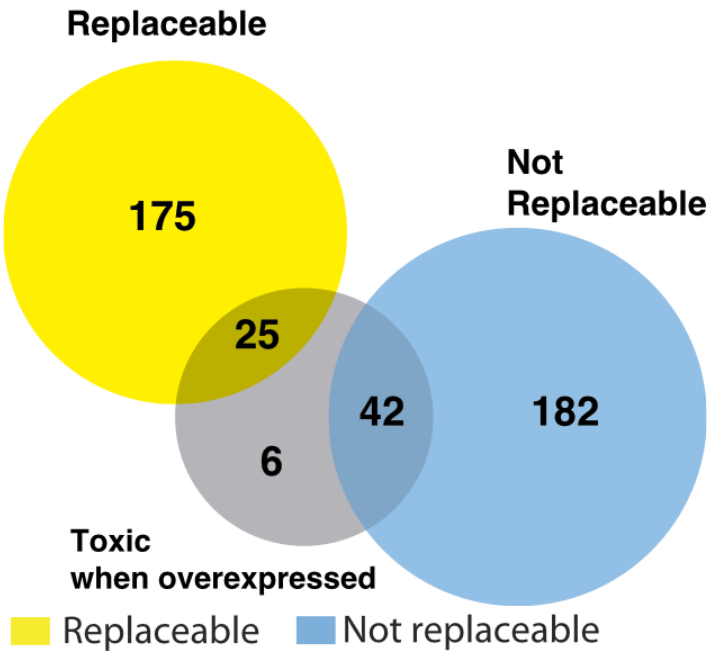


Figure 1.6.3. Substantial fraction of essential yeast genes can be replaced by their human counterparts.

On assaying >400 essential yeast genes, Kachroo *et al.* found that 200 (~47%) or nearly half of tested genes could be replaced by their human orthologs. Grey circle indicates gene that tended to be toxic when overexpressed some of which became replaceable when expressed appropriately. Figure adapted from Kachroo *et al*. (Kachroo *et al.* 2015).

Chapter 2: Humanization of yeast genes with multiple human orthologs[1]

A central question in evolutionary biology is how genes evolve and diverge from ancestral roles through the course of evolution. Previously, Aashiq Kachroo and Jon Laurent in the lab systematically addressed this question by performing functional cross-species gene swaps in *Saccharomyces cerevisiae* (Baker's yeast). These assays mainly tested the ability of human and in some cases even bacterial genes to replace their yeast counterparts. However, these tests were largely restricted to genes with no duplication in any lineage (Kachroo *et al.* 2015, 2017). They showed that ~47% of human and >60% of bacterial orthologs could actually substitute for their yeast equivalents.

Orthologous genes can however duplicate and expand in species-specific fashions to give rise to multiple gene orthology relationships between species and within gene families. This central question of how ancestral roles are distributed among orthologs (or co-orthologs) within gene families was not tested systematically and experimentally. This chapter describes the systematic humanization of essential yeast genes, one gene at a time, with their human orthologs that have undergone lineage-specific gene amplification. These assays allowed us to directly test which of the paralogous human genes have retained ancestral functionality and can replace the orthologous yeast gene. Similar to previous work, we take advantage of simple yeast growth assays to determine the functional divergence of genes in human multigene families. This chapter contains

[1]This chapter was published as Laurent, J.M., Garge, R.K., Teufel, A.I., Wilke, C.O., Kachroo, A.H., and Marcotte, E.M. (2020). Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. *PLOS Biology* 18, e3000627. This work was led by Jon M. Laurent and Aashiq H. Kachroo. My contributions were performing a subset of humanization experiments and helping with writing of the manuscript.

the results from >400 additional yeast humanization experiments bringing the overall tally of human genes known to functionally replace essential yeast genes to 280. Most interestingly, we observe a striking pattern of replaceability across different ortholog groups, with an obvious trend towards differential replaceability inside gene families, rarely observing replaceability by all members of a family.

Subsequently, we quantitatively measured many properties of the genes to discover rules governing the observed patterns of replaceability. This work represents one of the most extensive and rigorous direct functional studies of duplicated genes yet performed and extends the humanization assays to nearly all conserved, essential genes in yeast. Thus, contributing a significant number of new reagent strains that can be used by the community to study various aspects of human protein function in a simpler eukaryotic context.

2.1. ABSTRACT

Despite over a billion years of evolutionary divergence, several thousand human genes possess clearly identifiable orthologs in yeast, and many have undergone lineage-specific duplications in one or both lineages. These duplicated genes may have been free to diverge in function since their expansion, and it is unclear how or at what rate ancestral functions are retained or partitioned among co-orthologs between species and within gene families. Thus, in order to investigate how ancestral functions are retained or lost post duplication, we systematically replaced hundreds of essential yeast genes with their human orthologs from gene families that have undergone lineage-specific duplications,

including those with single duplications (1 yeast gene to 2 human genes, 1:2) or higher-order expansions (1:>2) in the human lineage. We observe a variable pattern of replaceability across different ortholog classes, with an obvious trend toward differential replaceability inside gene families, and rarely observe replaceability by all members of a family. We quantify the ability of various properties of the orthologs to predict replaceability, showing that in the case of 1:2 orthologs, replaceability is predicted largely by the divergence and tissue-specific expression of the human co-orthologs, i.e., the human proteins that are less diverged from their yeast counterpart and more ubiquitously expressed across human tissues more often replace their single yeast ortholog. These trends were consistent with in silico simulations demonstrating that when only one ortholog can replace, it tends to be the least diverged of the pair. Replaceability of yeast genes having more than 2 human co-orthologs was marked by retention of orthologous interactions in functional or protein networks as well as by more ancestral subcellular localization. Overall, we performed >400 human gene replaceability assays, revealing 50 new human–yeast complementation pairs, thus opening up avenues to further functionally characterize these human genes in a simplified organismal context.

2.2. INTRODUCTION

Humans and budding yeast differ dramatically with respect to cell and tissue organization, metabolism, motility, and environment; nonetheless, yeast has remained an important eukaryotic model for informing and answering questions related to our own biology. Thousands of human genes have clearly identifiable homologs in yeast (Remm

et al. 2001) **(Fig. 2.6.1A**), all deriving from genes of a common opisthokont ancestor living approximately a billion years ago (Kumar *et al.* 2017). Many studies have related function and disruption of yeast genes to their human counterparts, relying on a strong tendency for ancestral function to be retained between homologous genes (i.e., for sequence similarity to imply functional similarity). A more rigorous test of ancestral function being retained involves exchanging genes between related species to test for their ability to replace loss of the homologous genes. Indeed, interspecies swaps have been used for decades to identify homologs with equivalent function even before sequence comparison was as routine as it is today (Lee and Nurse 1987; Elledge and Spottswood 1991). More recently, large-scale systematic studies have identified many human genes that can substitute for their yeast equivalents and sustain yeast growth (Kachroo *et al.* 2015; Hamza *et al.* 2015; Sun *et al.* 2016; Yang *et al.* 2017). These systematic studies of functional replacement of yeast genes by homologs from humans (Kachroo *et al.* 2015; Hamza *et al.* 2015; Sun *et al.* 2016; Yang *et al.* 2017) and bacteria (Kachroo *et al.* 2017) demonstrate the power of interspecies gene swaps to directly test functional divergence and identify properties that determine functional conservation across vast evolutionary distances.

Most (but not all) prior yeast humanization studies have considered genes that have not acquired observable duplications in either the human or yeast lineage, referred to as 1:1 orthologs; that is, one yeast gene has one human ortholog. Two homologous genes related by speciation are termed “orthologs” and are distinguished from those related by gene duplication, termed “paralogs” (Conant and Wolfe 2008; Innan and Kondrashov 2010). However, many human genes in particular belong to gene families

that have duplicated and diverged along the human lineage. Importantly, an expanded family of human genes may often correspond to a single gene on the yeast lineage (or vice versa). (For simplicity, we will generally refer to such cases as co-orthologs, although lineage-specific gene deletions can complicate such assignments.)

More broadly, gene duplication is regarded as a major contributor to the production of new genetic material (Ohno 1970). Because of the immediate functional redundancy and dosage increase created following duplication, a common fate of duplicated genes is loss of one functional copy, returning to the ancestral state (Lynch and Conery 2000). However, duplicate genes can also either retain their ancestral roles or diverge and adopt new functions (Conant and Wolfe 2008; Innan and Kondrashov 2010). Lineage-specific duplications also have a major practical impact on how homologous genes between species are identified and defined. How ancestral functions are partitioned, lost, or retained during these duplication events is a major topic of study for evolutionary biology (Mushegian and Koonin 1996; Koonin 2005; Studer and Robinson-Rechavi 2009; Nehrt *et al.* 2011; Gabaldón and Koonin 2013; Dunn *et al.* 2018). Thus, expanding systematic tests of functional replacement across these expanded gene families could help address questions of how functions are retained or lost following duplication, as well as provide researchers with many more opportunities to study human gene function in the simplified context of budding yeast.

In this work, we sought to directly assay functional divergence within expanded gene families by using yeast humanization. We identified all essential yeast genes in yeast–human ortholog groups (orthogroups) that have undergone expansions in the human and/or yeast lineage and systematically replaced the yeast orthologs with each of

their human co-orthologs, assaying functional replaceability by complementation of a lethal growth defect. We find that duplicated human genes tend to differentially replace their yeast ortholog, rarely observing broad replaceability across members of expanded human gene families. Further, we quantified the ability of several protein-, gene-, or ortholog-based properties to explain the differential ability of human co-orthologs to replace and further support our observations with *in silico* simulations of protein evolution post duplication.

Collectively, our results suggest that within paralogous human gene families, at least one gene generally tended to retain ancestral function well enough to replace a billion-year–diverged ortholog in a yeast cell. These resulting strains and observational data serve as important resources for addressing questions in evolutionary biology, as well as for genetic and biotechnology applications. For example, they will enable new yeast-based drug screens against human genes, guide pathway and genome engineering efforts, and enable tests of human disease allele functionality, among other applications, all in the much simpler eukaryotic context of the yeast cell (Laurent *et al.* 2016). Our data underscore the remarkable extent to which genes are functionally equivalent between humans and yeast, demonstrating the power of distant model organisms for studying human processes.

2.3. Materials And Methods

2.3.1. Identifying orthologs

Orthologs were calculated with a local installation of InParanoid (Remm *et al.* 2001), using UniProt proteomes for the two species (downloaded November 2014). InParanoid identifies orthogroups between two species by first performing an all-versus-all BLAST search between the two species to identify bidirectional best hits (BBH). Each proteome is then subjected to an all-versus-all BLAST against itself to identify within-species homologs. The BBH pairs are used as seed matches, and any within-species genes from the self-BLAST that are at least as close to the gene of interest as its BBH in the other species are added to the ortholog group and termed an "in-paralog" or co-ortholog. We used the "table" output of InParanoid to identify orthogroup classes as follows: those groups with one listed yeast gene and two human genes were dubbed one-to-two (1:2) orthologs, whereas those with one yeast gene and more than two human genes were identified as one-to-more-than-two (1:>2) orthologs. Together, these two sets make up the 1:M ortholog class.

2.3.2. ORFeome cloning

Human genes were obtained from the human ORFeome collection (Lamesch *et al.* 2007). The ORFeome comprises a collection of *Escherichia coli* strains, each harboring a plasmid encoding a single human gene in a Gateway "entry" vector. Sequences cloned in the entry vectors are flanked by attL sites. To create expression vectors, each human entry vector was isolated from *E. coli*, added to a Gateway LR reaction with a Gateway "destination" vector, and transformed into competent *E. coli* to obtain expression clones

(Liang *et al.* 2013). As ORFeome clones lack stop codons, we modified the Advanced Yeast Gateway kit (Alberti *et al.* 2007) pAG416-GPD-ccdB destination vector, which does not encode a stop codon immediately outside of the cloning region, resulting in a tail of approximately 60 amino acids being added to any protein expressed from it. We thus mutagenized the vector downstream of the cloning region to introduce a stop codon, shortening the tail to six amino acids (this plasmid is termed pAG416-GPD-ccdB+6Stop)(Kachroo *et al.* 2015). Entry and expression clones were verified by sequencing into the gene sequence from the upstream and downstream regions of the plasmid.

2.3.3.MGC cloning

For human genes not available in the ORFeome, we obtained clones from the Mammalian Gene Collection (Temple *et al.* 2009), a collection of sequence-verified human cDNA sequences. To obtain entry vectors for these genes, we designed primers for each gene that would amplify the coding sequence while adding attB sites to either end of the human gene and performed PCR using the MGC plasmid as a template. PCR products were combined with plasmid pDONR221 in a Gateway BP reaction (Liang *et al.* 2013) and transformed to *E. coli* to obtain entry clones. The entry clones were then combined in a Gateway LR reaction with p416-GPD-ccdB+6Stop and transformed to *E. coli* to obtain expression clones. Each entry and expression clone was verified by sequencing the clone boundaries at each end of the gene.

2.3.4. Functional replaceability assays

Yeast strains were grown in a 96-well format in YPD medium supplemented with G418 (200 µg/ml). The strains were transformed with matched expression clones or empty control vectors and selected on minimal medium lacking uracil. Complementation assays were performed as follows:

1. *TS assays*.

Each of the strains in the TS strain collection (Li *et al.* 2011) encode a yeast protein with a mutation that allows growth at permissive temperatures of 22–26 °C but not at the restrictive temperature of 35–37 °C. We therefore identified human genes capable of rescuing growth of the mutant at restrictive temperature on selective plates. Each strain was transformed and assayed separately with either the human gene-expressing vector or the empty vector control by growing transformed strains in the following manner: −Ura dextrose medium at the permissive temperature (26 °C), which serves as a control for transformation efficiency and/or toxicity because both the yeast and the human gene are expressed; and −Ura dextrose medium at the restrictive temperature (37 °C), testing for human gene functional replacement under conditions in which the corresponding yeast gene is nonfunctional.

2. *hetKO assays.*

Strains in the hetKO collection (obtained from ATCC) are heterozygous diploid strains, each harboring one allele of a yeast gene knocked out by replacement with the KanMX kanamycin-resistance cassette, allowing for selection on G418 (Pan *et al.* 2004). We transformed human expression clones or an empty control vector into appropriate strains and selected on −Ura G418 medium in a 96-well format. Transformants were then

replated on GNA-rich presporulation medium containing G418 and 50 mg/L histidine. Individual colonies were then inoculated in liquid sporulation medium containing 0.1% potassium acetate, 0.005% zinc acetate, and incubated with vigorous shaking at 26 °C for 3–5 days, after which sporulation efficiency was estimated by microscopy, and the mixture was then resuspended in water and equally plated on two assay conditions: (1) −G418 Magic Marker dextrose medium (−His −Arg −Leu +Can −Ura) incubated at 30 °C. The haploid spores that carry the wild-type allele grow in this medium, providing us with the control for sporulation efficiency. This condition also assays for toxicity of the human gene if the haploid spores fail to grow. (2) +G418 Magic Marker dextrose medium (−His −Arg −Leu +Can −Ura) containing 200 μg/ml G418. In the absence of the human gene (as for control transformants), the resulting haploid knockout strain is expected not to grow, providing an assay for replaceability in strains expressing the human gene. Cases with approximately equal numbers of colonies growing in the absence or presence of G418 were considered functional replacements. For cases with ambiguous growth (marked by moderate numbers of isolated colonies growing on the +G418 medium relative to −G418 medium), we screened varying quantities of the sporulation mixtures.

Positive assays were verified independently. Individual colonies were isolated from selective plates and were assayed for growth defects on YPD or Magic Marker medium + G418 (**Fig. 2.6.1B, S4 and S5**). After growth on YPD + G418, each strain was spotted on 5-FOA agar to test plasmid dependency.

3. ***Tetrad dissection and plasmid loss assays***

For human gene replaceability assays performed in the yeast hetKO collection that were ambiguous in our large-scale screen, we performed tetrad dissections to more clearly test for complementation (**Fig. 2.6.S2**). In total, 33 human genes were assayed and analyzed. We transformed each human expression clone or empty vector control into the appropriate yeast strains and selected on SC-Ura + G418 (200 μg/ml) to select for the human gene expression vector (CEN, Ura+) and yeast gene knockout (KanMX marker) simultaneously. Transformants were then plated on GNA-rich presporulation medium containing G418 (200 μg/ml). Individually isolated colonies were inoculated into liquid sporulation medium containing 0.1% potassium acetate, 0.005% zinc acetate, and were incubated with vigorous shaking at 25 °C for 3–5 days. Following this, sporulation efficiency was estimated by microscopy, and successful sporulations were subjected to tetrad analysis. In all, 15–20 μL of each sporulation was digested with an equal volume of Zymolyase (5 mg/ml stock) for 30–45 minutes to remove the ascus coats. The digestions were diluted 1:1 with sterile water, after which 20–30 μL of the Zymolyase-treated spores was carefully applied to a tilted YPD plate using a pipette, allowing the droplet of cell suspension to gently run down the agar surface. The plates were dried and visualized on the tetrad dissection microscope. For each human gene, a minimum of five tetrads were dissected. Dissections were selected and replica plated both on 5-FOA (for plasmid counterselection) and YPD + G418 (for yeast null allele selection). A successful complementation consists of 2:2 segregation with survival on YPD + G418 and failure to grow on 5-FOA **(Fig. 2.6.S2, Table 2.7.1)**. We subsequently performed quantitative growth assays (in triplicate) on tetrads passing the 5-FOA/G418 segregation test. Each humanized tetrad was grown in three different media conditions: YPD, YPD + G418, and

SC-Ura. Subsequently, dissections were replica plated to SC-Ura to orthogonally confirm plasmid dependencies in the G418-resistant clones. Each medium-specific growth profile (shown in one of the conditions as in **Fig. S4 and S5**) was analyzed and quantified to detect any growth defects in yeast. In all, 14 out of 33 human genes assayed in this manner showed functional replaceability of the yeast gene function **(Table 2.7.1)**.

2.3.5. Isolating haploid humanized yeast strains and quantitative growth curve assays

Select yeast hetKO strains carrying human gene expression vectors (CEN, Ura+) showing functional replaceability were sporulated and plated on independent petri plates to obtain single colonies. Each colony was tested for plasmid loss in the presence of 5-FOA. Colonies that did not grow in the presence of 5-FOA were subjected to further analysis to quantitatively measure their growth rates. Yeast strains were either precultured in liquid YPD + G418 (200µg/ml) or −Ura dextrose selective medium + G418 (200 µg/ml) for 2 hours or overnight, respectively. Each culture was diluted in YPD or −Ura dextrose medium to an OD of approximately 0.1 in 100 or 150 µL total volume in a 96-well plate. Plates were incubated in a Synergy H1 shaking incubating spectrophotometer (BioTek), measuring the OD every 15 minutes over 48 hours. Growth curves were performed in triplicate for each strain by splitting the preculture into three independent cultures for each 48- to 60-hour time course (**Fig. 2.6.S4A, S5A and S5C**).

In the case of TS humanized yeast strains, the growth assays were performed first at permissive temperatures (25–26 °C) for a 48- to 60-hour time course. These growth assays were largely identical to the empty vector transformed yeast strains. The strains were then shifted to a restrictive temperature of 37 °C, and we similarly repeated the

growth assay for a 48- to 60-hour time course (**Fig. 2.6.S4B and S5B**). Growth curves were performed in triplicate for each strain by inoculating cells from the same preculture into three independent cultures.

For computational analyses of the trends underlying replaceability, we computed a diverse set of features **(Table 2.7.3)**, based in part on those from Kachroo and colleagues (Kachroo *et al.* 2015), as follows:

2.3.6. Sequence properties

Sequence features for human genes obtained from the human ORFeome (Lamesch *et al.* 2007) were calculated using the ORFeome-provided fasta file for the sequences, which was downloaded from http://horfdb.dfci.harvard.edu/hv7/docs/human_orfeome71.tar.gz. For clones not obtained from the human ORFeome collection, we calculated sequence features using the longest annotated transcript or its translation from Ensembl version 74, available at http://Dec2013.archive.ensembl.org/index.html. To analyze protein features, we translated the above nucleotide sequence for each gene using the standard genetic code. The following protein sequence features were considered:

Sequence length:

Sc_Length

Hs_Length

Sc-Hs_LengthDifference

Sc-Hs_AbsLengthDifference

Length was calculated as the count of amino acids in the protein. LengthDifference was calculated as the length of the human ortholog protein subtracted from the length of the yeast ortholog. AbsLengthDifference is the absolute value of LengthDifference.

Sequence similarity:

Sc-Hs_PercentIDLongest

Sc-Hs_PercentIDAligned

Sc-Hs_PercentSimLongest

Sc-Hs_PercentSimAligned

Hs_Orthoscore

Hs_OrthoRank

Orthologous pairs were first identified by InParanoid (Remm *et al.* 2001). Global alignments for all preidentified ortholog pairs were then calculated using NWalign (http://zhanglab.ccmb.med.umich.edu/NW-align/) with BLOSUM62 and gap open penalty of −11 and extension −1. Identity and similarity were calculated as the fraction of identical amino acids and amino acids with positive BLOSUM score in the alignment, respectively. “Longest” refers to amino acid identity or similarity calculated as a fraction of the longer of the two orthologs. “Aligned” refers to calculating identity or similarity calculated as a fraction of the length of the aligned region of the sequences. OrthoScore and OrthoRank refer to the scores assigned by InParanoid. For each orthogroup, InParanoid calculates a confidence score between 0 and 1 for each in-paralog that represents how similar it is to the seed ortholog. Rank is simply a ranked ordering of the

in-paralogs of a group based on their orthoscore, with 1 being the LDO ortholog, and higher values being further away from the seed.

Codon usage:

Sc_CAI

Sc_CBI

Sc_FOP

Hs_CAI

Hs_CBI

Hs_FOP

Calculated from the amino acid sequences using CodonW (http://sourceforge.net/projects/codonw/). Properties for the human genes were calculated using the yeast codon optimality table as a measure of closeness or divergence from yeast optimality.

2.3.7. Network properties

Network features were calculated using custom Python scripts, typically utilizing the package "networkx" (available from http://networkx.github.io/documentation/latest/download.html). When applicable (e.g., HumanNet, YeastNet), provided weights of interactions were taken into account when calculating these features. Otherwise, a default weight of 1.0 was used. Network features were defined as follows: Degree represents the count of interaction partners for a node in a given network. Betweenness represents network centrality, a measure of how central in

a network a given node is, calculated as the number of shortest paths between all node pairs in a network that pass through a given node. Clustering represents the node clustering coefficient, calculated as the fraction of edges that could possibly be present in a node's neighborhood that are actually present. FractionComplementing is the fraction of interaction partners observed to complement (including results obtained in our 1:1 humanization assays (Kachroo *et al.* 2015)). FractionOrthologPartners is the fraction of a gene's interaction partners that have orthologs in the other species and interact with the gene of interests' ortholog in the corresponding network (i.e., if gene Sc-*A* interacts with Sc-*X*, Sc-*Y*, and Sc-*Z*, and Hs-*A* interacts with Hs-*X* and Hs-*Y* but not Hs-*Z* [which is a legitimate gene], the FractionOrthologPartners for Sc-*A* is 2/3 or 0.66). It represents a measure of the degree to which interactions are maintained between orthologs in the two species.

BIOGRID:

[Hs|Sc]_BIOGRID_[Degree|Betweenness|Clustering|FractionComplementing|FractionOrthologPartners]

Calculated from interactions present in BIOGRID 3.1.93 (Stark *et al.* 2006) , using only those interactions annotated as "physical interactions."

Functional networks:

[Hs|Sc]_*Net_[Degree|Betweenness|Clustering|FractionComplementing|FractionOrthologPartners]

Human and yeast functional gene network features were calculated based on HumanNet (Lee *et al.* 2011) and YeastNet (McGary *et al.* 2007), respectively. The final

sum log-likelihood score reported for each interaction was employed as an edge weight for calculations.

Abundance properties:

[Hs|Sc]_ProteinAbundance

[Hs|Sc]_TranscriptAbundance

[Hs|Sc]_RPFAbundance

[Hs|Sc]_TranslationEfficiency

ScHs_SubCellHamming

Protein abundances were used as reported by Kulak and colleagues (Kulak *et al.* 2014). TranscriptAbundance, Ribosome Protected Fragments (RPF), Abundance, and TranslationEfficiency were calculated from Guo and colleagues (human) (Guo *et al.* 2010) and Ingolia and colleagues (yeast) (Ingolia *et al.* 2009). Translation efficiency was calculated as the ratio of RPF reads to mRNA reads. SubCellHamming is a measure of the difference in subcellular localization between a yeast–human ortholog pair, calculated using data from the COMPARTMENTS database ((Binder *et al.* 2014), downloaded in February 2017). We utilized the "benchmark" sets to create, for each protein in its respective species, a binary vector of subcellular localization across 11 compartments (cytoskeleton, cytosol, endoplasmic reticulum, extracellular space, Golgi apparatus, lysosome, mitochondrion, nucleus, peroxisome, plasma membrane) that were common to both human and yeast in the database. Then, for each ortholog pair, we calculated the hamming distance between their respective subcellular localization vectors as the SubCellHamming value.

2.3.8. Calculating predictive strength of features

The predictive power of each feature was calculated as the AUC when treating that feature as an individual classifier. Each feature was sorted in both ascending and descending directions, retaining the direction providing an AUC > 0.5. To assess significance, a permutation procedure was performed as follows: for each feature, the replaceable/nonreplaceable status of each ortholog pair was shuffled (retaining the original ratio of replaceable-to-nonreplaceable assignments), and the AUC was calculated. The shuffling procedure was carried out 1,000 times for each feature, and the mean AUC values and their standard deviations were reported.

1.*Features of expanded orthogroups.* In order to avoid overweighting expanded gene families and compensate for uneven sampling, we considered median properties of genes in each orthogroup as follows.

For each orthogroup, we collapsed all yeast–human ortholog pairs with the same status (replaceable or not) into a single case, with the value of each feature calculated as the median value of the collapsed ortholog pairs for that feature from the original table. Thus, each 1:M orthogroup is collapsed to be represented by either two pairs (complement and noncomplement) or one pair (either complement or noncomplement).

2.3.9. Simulations

We constructed a simulation of protein evolution in which one member of an interacting pair was duplicated. The simulation was initialized with the yeast SMT3–UBC9 complex, a small heterodimeric protein complex, as the resident genotype (PDB: 2EKE) (Duda *et al.* 2007). This complex initially had two subunits, which we refer to as A and B. We duplicated the B subunit and refer to it as B′. Our simulation protocol and

setup (Teufel *et al.* 2019) is based on an accelerated origin-fixation model (Kachroo *et al.* 2015; Teufel and Wilke 2017). Here, we further analyzed five of the previously published sets of 100 simulation trajectories (Teufel *et al.* 2019) (summarized in the **Fig. 2.6.6A Selection Schemes panel**). Briefly, under the first selection scheme, we enforced selection for A to bind B and for A to bind B′ (bind both). In the second scheme, selection acts on the ability of A to bind B or A to bind B′, and the maximum stability of those interactions was considered (bind max). In the third scheme, the ability of A to bind B was selected for, but the ability of A to bind B′ was not (bind B). In the fourth scheme, the ability of A to bind B was also selected for, but the ability of A to bind B′ was selected against (bind B and not B′). We also implemented a control selection scheme in which selection does not act on the ability of A to bind either B or B′ (no bind). We performed 100 replicates of each of these selection schemes.

The percent of simulations in which an evolved duplicate has the ability to bind the ancestral partner **(Fig. 2.6.6A and 6B)** corresponds to Figs S7A-B in Teufel and colleagues' publication (Teufel *et al.* 2019). We further analyzed these data by examining instances in which only one of the duplicates is able to bind the ancestral partner at the end of the simulation run. We recorded each duplicate's divergence, defined as the fraction of amino acid positions that were nonidentical between the initial and final sequences of a simulation run **(Fig. 2.6.6C)**. For each selection scheme, the distribution of the divergence of duplicates that can bind the ancestral partner was compared with the distribution of divergence of duplicates that cannot bind the ancestral partner with a paired *t* test.

Simulation data and scripts are available at https://github.com/a-teufel/Laurent_etal_2020.

2.4. Results and Discussion

2.4.1. Identifying and selecting orthologs in expanded orthogroups and ortholog replaceability assays

Previous work from our group demonstrated that roughly half (47%) of one-to-one (1:1) yeast-to-human orthologs were capable of replacing essential yeast genes across a panel of three complementation assays **(Fig. 2.6.1A)**(Kachroo *et al.* 2015). We had specifically restricted these earlier tests to ortholog groups for which no lineage-specific gene duplications were easily identified to mitigate any effects of functional redundancy between paralogs. In the present study, we now focus on these cases of ortholog groups that have undergone lineage-specific gene family amplifications **(Fig. 2.6.1A)**. We again restricted our test set to include only those groups in which the yeast gene was annotated as essential for growth under standard laboratory conditions **(Fig. 2.6.1A, Table 2.7.1)** (Winzeler *et al.* 1999; Tong *et al.* 2001) and tested each human co-ortholog for its ability to individually complement loss of its yeast ortholog. We tested all human gene:yeast strain combinations for which reagents were available, even when not all members of an orthogroup could be tested.

In total, we identified 2,073 orthogroups, comprising 4,556 ortholog pairs (2,424 yeast proteins and 3,690 human proteins), using InParanoid (Remm *et al.* 2001). Considering only yeast proteins essential for growth in standard laboratory conditions,

there are 1,001 ortholog pairs **(Fig. 2.6.1A)** distributed in 706 orthogroups. In our previous study, we obtained results for 424 human genes belonging to 1:1 orthogroups **(Fig. 2.6.1A)** (Kachroo *et al.* 2015). The remaining orthogroups comprise 468 additional pairs with essential yeast orthologs **(Fig. 2.6.1A)** and are the focus of this study. These expanded protein families were further split according to whether they had expansions only on the human lineage (one-to-many [1:M]) or expansions on either the yeast or both lineages (many-to-one or many-to-many, collectively referred to here as M:M), resulting in 378 1:M pairs (involving 140 orthologous yeast genes) and 90 M:M pairs (representing 36 yeast genes and 83 human genes) **(Fig. 2.6.1A)**. The 1:M orthogroups varied widely in size, with the majority having only two human members, and the largest orthogroup (the melanoma antigen gene [MAGE] family) comprising 28 human co-orthologs **(Fig. 2.6.S1)**.

We obtained human gene clones for our assays from either the human ORFeome collection or human sequence-verified Mammalian Gene Collection (Lamesch *et al.* 2007; Temple *et al.* 2009). Human genes were subcloned via Gateway cloning (Alberti *et al.* 2007) into yeast expression vectors under the control of the yeast GPD promoter, driving constitutive, robust expression. We performed complementation assays in one or both of two distinct yeast strain backgrounds: The first, referred to as the heterozygous diploid knockout deletion (hetKO) collection, represents a collection of diploid yeast strains, each of which contains a hetKO of a single gene, with a selectable marker cassette allowing isolation of the haploid knockout post sporulation (Pan *et al.* 2004). The second mutant background is a library of temperature-sensitive (TS) haploid yeast strains, each harboring a mutant allele encoding a TS form of the protein (Li *et al.* 2011) that can

be inactivated by growth at restrictive temperatures (typically 36 or 37 °C). These two strain backgrounds allow precise conditional loss of an essential yeast gene or protein for assaying complementation by its human ortholog(s) **(Fig. 2.6.1B, Materials and Methods)** (We had previously demonstrated that 41 of 42 assayed strains could be rescued by plasmid-based expression of the corresponding yeast gene (Kachroo *et al.* 2015, 2017), confirming the utility of this approach.) We verified our assay results by isolating haploid humanized yeast gene knockout strains (either using Magic Marker medium or by tetrad dissection) while simultaneously verifying dependency on the human gene-encoding plasmid **(Fig 2.6.S3, , Materials and methods)**. We further performed quantitative growth assays for each of the humanized yeast strains to more accurately characterize the robustness of complementation. The majority of complementing human genes exhibits similar growth profiles to the parental yeast strain **(Fig. 2.6.1B, S4 and S5)**.

In total, we obtained informative results for 308 of the 378 1:M ortholog pairs. ("Informative" here refers to those assays in which all controls behaved appropriately. See **Materials and methods** for details.) This translates to successful results from at least one human ortholog for approximately 93% (130 of 140) of testable 1:M yeast genes. Of the 90 M:M pairs, 29 were successfully assayed. **(Fig. 2.6.1A, Table 2.7.1)**. Because of the scarcity of replaceable M:M ortholog pairs and the complicated nature of their orthology relationships, we simply report the results of their assays but do not include them in any subsequent analysis.

2.4.2. Orthologs in expanded human gene families differentially replace their yeast orthologs

Of the 130 successfully tested 1:M yeast genes, 52 (40%) had at least one human ortholog that replaced, whereas 78 were not replaceable by any tested human ortholog **(Fig. 2.6.2, S4 and S5, Table 2.7.1)**. Notably, this rate of replaceability from the yeast gene perspective is similar to that previously observed for 1:1 orthologs (40% compared with 47%) (Kachroo *et al.* 2015).

From the perspective of human genes in these expanded gene families, 74 of the 308 successfully tested 1:M human genes could functionally replace their yeast ortholog, whereas 234 could not **(Fig. 2.6.2A, S4 and S5)**. Of the orthogroups with at least one complementing human gene, the majority (34/52 groups, or 65%) showed differential replaceability, whereas all human co-orthologs in a group replaced only in a few cases (12/52 or approximately 23%) **(Fig. 2.6.2C, S4 and S5, Table 2.7.1)**. In marked contrast, the great majority of human genes belonging to the M:M class were not able to replace their yeast orthologs, with only three human genes replacing in these cases, distributed across three different ortholog groups **(Figs 2.6.2C, S4C, Table 2.7.1)**.

When available, we compared our 1:M and M:M replaceability results with published reports, observing a strong agreement, with our assays recapitulating literature results (here, defined solely by the YeastMine database (Balakrishnan *et al.* 2012)) with 88% accuracy (**Table 2.7.2**). Accounting for our previous assays as well as other published tests, we report an additional 50 novel yeast–human complementation pairs. With previously published cases from the 1:1 set, this brings our total count of known yeast genes with swappable human orthologs to 280 (Kachroo *et al.* 2015)**(Fig. 2.6.2B)**.

2.4.3. Computational analysis of trends governing replaceability

Because the ability of human co-orthologs to functionally replace their singleton yeast orthologs was generally differential within expanded gene families, we sought to identify characteristic features governing selective replaceability of specific human genes in expanded orthogroups. To that end, we assembled and/or calculated a number of quantitative properties for all genes and ortholog pairs. These properties include direct properties of the genes in an orthogroup (e.g., protein length, codon usage bias), as well as comparative properties between co-orthologs (e.g., protein sequence similarity, length difference). We could then assess each property for its ability to explain replaceability as the area under a receiver–operator characteristic (ROC) curve, treating each feature as an individual classifier, in a similar manner to the previous 1:1 ortholog study (Kachroo *et al.* 2015). To simplify pairwise comparisons, we considered median values of co-orthologs within each orthogroup that had the same replaceability status (i.e., in which both could replace or not) **(Fig. 2.6.S6)**. Our analysis is robust to missing data, ignoring features for gene pairs that had no informative assay. Of the 378 testable 1:M pairs, the majority (208 pairs in 104 orthogroups) belonged to a group that had only two human co-orthologs (referred to as the 1:2 class) **(Fig. 2.6.S1)**. The remaining orthogroups with more than two human co-ortholog members were dubbed the 1:>2 class. Owing to this disparity, we chose to consider these two groups separately in subsequent analyses.

2.4.4. Differential replaceability in 1:2 orthogroups is predicted by co-ortholog divergence and mRNA expression specificity

In the 104 1:2 ortholog groups, we obtained informative results for 171/208 human genes (94/104 yeast genes) **(Fig. 2.6.S1 and S4, Table 2.7.1)**. Forty-five of the

tested human genes functionally replaced the yeast gene, whereas 127 could not (36 human genes were either not tested or did not result in an informative assay). Seventy-seven of these groups were completely tested (i.e., both human co-orthologs were assayed and yielded informative results), 17 groups had only a single tested human gene, and 10 had no informative assay results (**Table 2.7.1**). From the yeast genes' perspectives, 34/95 (36%) of those assayed were replaceable by at least one human ortholog.

Applying our computational analysis pipeline to the 1:2 orthogroups revealed several explanatory features. Most prominent in its explanatory ability, and in opposition to results from assays of 1:1 orthologs (Kachroo *et al.* 2015), was the relative divergence of the human co-orthologs from each other and from the yeast ortholog **(Fig. 2.6.3A)**. In particular, the highest predictive power (measured as area under the ROC curve [AUC]) was seen for the InParanoid ortholog rank (Hs_OrthoRank), which ranks the human co-orthologs from more diverged (MDO) to least diverged (LDO) from the yeast counterpart, and ortholog score (Hs_OrthoScore, a score ranging from 0 to 1, with scores closer to 1 meaning nearer to the yeast ortholog) **(Fig. 2.6.3A)**. Of the 77 completely tested 1:2 orthogroups, there were 23 groups in which the two human genes displayed differential ability to replace (i.e., one human gene replaced, whereas the other did not). These groups are particularly interesting because they provide an opportunity to investigate properties that distinguish co-orthologs that can replace the yeast gene or not within the same orthogroup. To assess this difference, we performed the ROC analysis for each feature specifically on the set of differentially replaceable 1:2 orthogroups **(Fig. 2.6.3B)**. In this restricted set, the relative co-ortholog divergence strengthened its

standing as the most predictive feature in these orthogroups, demonstrating that the LDO of two human co-orthologs was considerably more likely to replace **(Fig. 2.6.3B)**. Indeed, in this differentially replaceable set, 86% of the replaceable human genes are also the LDO in their respective orthogroups (a trend similar to that observed by Hamza and colleagues (Hamza *et al.* 2015)). Further, when we compared the divergence of the differentially replaceable set ("one replaces") with the "both replace" and "none replace" sets, the average InParanoid ortholog scores of the MDO were similarly low for "one replaces" and "none replace" (0.52 and 0.47, respectively), whereas the MDO in the "both replace" class is notably more similar to the LDO (0.73) **(Fig. 2.6.4A)**. The mean sequence identities between human and yeast orthologs for the "both replace" and "one replaces" classes are not significantly different, whereas the "none replace" class has a modestly lower mean identity to the yeast ortholog **(Fig. 2.6.4B)**. Thus, although human co-orthologs in the three complementation classes have diverged from their yeast ortholog to a similar extent, those in the "both replace" class have maintained similarity to each other and are thus more likely to both replace **(Fig. 2.6.4C-4E)**.

An additional significantly predictive feature appearing in the differentially replaceable 1:2 class pertains to the tissue-specific expression of the human orthologs. Specifically, co-orthologs with more widespread (less–tissue-specific) expression were more likely to replace **(Fig. 2.6.3B)**. This feature is particularly predictive of the differentially replaceable set but was not a strong feature for the full 1:2 orthogroup results **(Fig. 2.6.3A)**.

2.4.5. MDO yet replaceable 1:2 orthologs are highly expressed and more similar to their yeast counterparts

Despite the fact that in the majority of the 1:2 cases the LDO human co-ortholog replaced the corresponding yeast gene, there were several cases wherein the MDO co-ortholog was replaceable. Of 80 informative assays for human MDOs, 12 replaced the yeast ortholog. To determine what drives these replacements, we applied our ROC analysis scheme to the set, not restricting it to completely assayed groups or those in which the LDO also replaced. For these genes, the ability to complement is dominated by protein sequence similarity to the yeast ortholog as well as protein abundance features **(Fig. 2.6.3C)**. These results are in line with the previous observation that human co-orthologs that both replace are more similar than those in which only one or none replaces, but they also suggest that these highly diverged co-orthologs retain high expression levels. Thus, some highly diverged human co-orthologs still seem to maintain ancestral functionality, irrespective of the other co-ortholog's function.

2.4.6. Human co-ortholog replaceability in 1:>2 orthogroups is marked by conserved interactions and subcellular localization

In the case of 1:>2 orthogroups, 137/170 human genes were successfully assayed, 30 of which replaced the yeast ortholog, whereas 107 did not (22% replaceable). There were 23 1:>2 groups completely assayed for all human co-orthologs, represented by 78 pairwise tests **(Fig. 2.7S5A and S4B, Table 2.7.1)**. From the yeast perspective, 18 of 36 yeast genes with one or more successful assays were replaceable by at least one human co-ortholog. Six of those were replaceable by only one co-ortholog, and only two yeast genes in the 1:>2 set were replaceable by all of their human co-orthologs **(Table 2.7.1)**.

Because orthogroups with more than two human members had differing rates of replaceability within them and not all human genes could be assayed, we again collapsed the orthogroups using median values of replaceable versus nonreplaceable genes in each 1:>2 orthogroup prior to analysis **(Fig. 2.6.S6)**. Unlike for 1:2 orthogroups, but reminiscent of prior 1:1 findings, although sequence similarity appears near the most explanatory features, the ability of 1:>2 orthologs to replace was most strongly marked by functional context **(Fig. 2.6.5A)**. In particular, the dominant predictive features included the fraction of conserved protein–protein interactions along with their centrality in their respective functional interaction networks and the number of interactions in those networks, albeit to a lesser extent. Specifically, human co-orthologs that have maintained a higher fraction of orthologous interaction partners relative to the yeast ortholog were more likely to replace. Further, human co-orthologs with relatively higher centrality in a functional network were more likely to be replaceable. Thus, in the case of highly expanded gene families, paralogs that maintain ancestral protein contacts and centrality in their interaction networks tend to be more replaceable **(Fig. 2.6.5A)**, consistent with previous suggestions that "functional orthologs" retain a higher proportion of shared network contact (Bandyopadhyay *et al.* 2006) and observations that centrally placed proteins in interaction networks tend to carry out crucial cellular functions (Zotenko *et al.* 2008) that have been retained over vast evolutionary timescales.

We further sought to identify features that distinguished replaceable co-orthologs from nonreplaceable ones within the same orthogroup for the 1:>2 set. We again considered median features of both replaceable versus nonreplaceable co-orthologs within those 1:>2 orthogroups that showed differential replaceability in a manner similar

to the differential 1:2 set. No features showed an AUC more than 2 standard deviations above the mean of permutation tests, likely due to the small size of this specific orthogroup set. Nonetheless, the strongest trend observed was for replaceable co-orthologs to be localized in more similar subcellular compartments (Binder *et al.* 2014) in human cells as the yeast ortholog(s) **(Fig. 2.6.5B)**. Our observations are consistent with a model that at least one co-ortholog in an expanded human gene family will tend to retain essential ancestral functionality by maintaining ancestral interactions and network centrality, as well as similar cellular localization **(Fig. 2.6.5C)**.

2.4.7. Simulations suggest MDO duplicates are less likely to bind their ancestral interaction partners

To assess whether these trends would be evident in a controlled evolution experiment, we performed in silico simulations of functional divergence in a duplicated gene family. We analyzed a small heterodimeric protein complex consisting of A and B subunits (the SMT3–UBC9 protein complex (Duda *et al.* 2007; Kachroo *et al.* 2015)) in which the subunit B was duplicated in silico to yield A, B, and B′. Using binding of A to B and/or B′ as a proxy for functionality, we carried out evolutionary simulations of molecular structural divergence (considering all atom models using the Rosetta molecular modeling suite (Rohl *et al.* 2004; Leaver-Fay *et al.* 2011)). As described previously (Teufel *et al.* 2019), we examined functional replaceability across five different selection scenarios. All of the simulations assumed that selection acts on the stability of each subunit but differed in how they imposed selective pressure on binding. We ran 100 replicates for each of the five selection schemes and quantified replaceability of each of the subunits (measured as continued ability to bind to the ancestral partner) over

sequence divergence **(Fig. 2.6.6A and 6B).** Notably, selection for A to bind both B and B′ results in a continued ability for either B or B′ to bind the ancestral variant A, whereas application of diversifying selection to prevent binding to B′ results in a rapid decay in the ability of B′ to bind A **(Fig. 2.6.6B)**.

We then looked across our simulated lineages at those cases in which one of the duplicates functionally replaces (i.e., binds to the ancestor) and the other does not. We found a systematic pattern across all five selection schemes that the nonbinding duplicate tends to be the MDO one **(Fig. 2.6.6C**; all pairwise comparisons within selection schemes are significant, $p < 0.001$ paired t test**)**. These results mirror our experimental humanization findings for the relative divergence of 1:2 human co-orthologs that the replaceable duplicate tends to be less diverged than the nonreplaceable duplicate **(Fig. 2.6.4)**.

2.5. CONCLUSIONS

By extending the scope of our systematic yeast humanization assays to include those yeast genes that have more than one human ortholog, we successfully added 337 human genes to our tested set **(Fig. 2.6.2)**. We have therefore greatly expanded the set of human genes that can now be functionally studied in the simplified unicellular eukaryotic context of budding yeast by adding 50 novel human genes to those that can successfully replace their yeast ortholog. Overall, we found that yeast genes with duplicated human orthologs can be replaced by at least one co-ortholog at a slightly lower rate (40%) than 1:1 orthologs (Kachroo *et al.* 2015). Of those that could be replaced, the clear majority

was replaceable only by one or two co-orthologs, rather than being broadly replaceable by many human genes in the same family. This pattern of replaceability between and within orthogroups suggests divergence among gene family members away from ancestral functions, at least to the degree that yeast complementation assays can probe this effect (Koonin 2005; Gabaldón and Koonin 2013).

After testing many properties of the gene families for their ability to explain replaceability, our analysis revealed divergent patterns across groups that have two (1:2) or more than two (1:>2) human co-ortholog members. In the case of the 1:2 class, the top predictors were dominated by features that capture divergence from the yeast ortholog; in particular, we observed that the less-diverged variant strongly tended to be the replaceable one. This observation was supported by computational simulations of protein divergence after duplication, in which the LDO of the duplicates retained ancestral binding ability in most cases. In addition to sequence divergence, we also observed a strong trend of replaceable 1:2 human co-orthologs to have broader, less–tissue-specific expression. Somewhat rarely, the MDO human co-ortholog could replace, and these tended to be more similar to their yeast counterpart and expressed at a higher level, perhaps due to retaining important ancestral functions. In the case of the 1:>2 set, replaceability was marked largely by network-based properties of the genes. Replaceable co-orthologs in this set seem to have retained more ancestral interaction partners as well as higher centrality, even across a billion years of divergent evolution, likely indicative of their functional importance. Although not significant by our criteria, we also observed replaceable 1:>2 co-orthologs being expressed in similar subcellular compartments to

their corresponding yeast gene, highlighting a trend to maintain their ancestral localization.

Overall, our extended set of humanization assays and their analysis reveals trends in functional divergence among co-orthologs. We observed a strong tendency for orthogroups to exhibit only one or a few swappable human genes rather than many. We also extended our observation that network centrality and interaction properties aid in determining how ancestral gene function across orthologs is retained over deep evolutionary timescales. Our study was limited to those yeast genes that are known to be essential for growth, providing us with a straightforward phenotypic readout for our systematic screening strategy. Future studies incorporating custom phenotyping will be needed to determine whether similar rates and trends of replaceability are obeyed by nonessential genes.

Such assays help to advance our understanding of duplicate gene evolution in the billion years since yeast and humans diverged (Douzery *et al.* 2004) and add powerful reagents to study myriad human processes and develop therapies in a simpler eukaryotic surrogate.

2.6. FIGURES

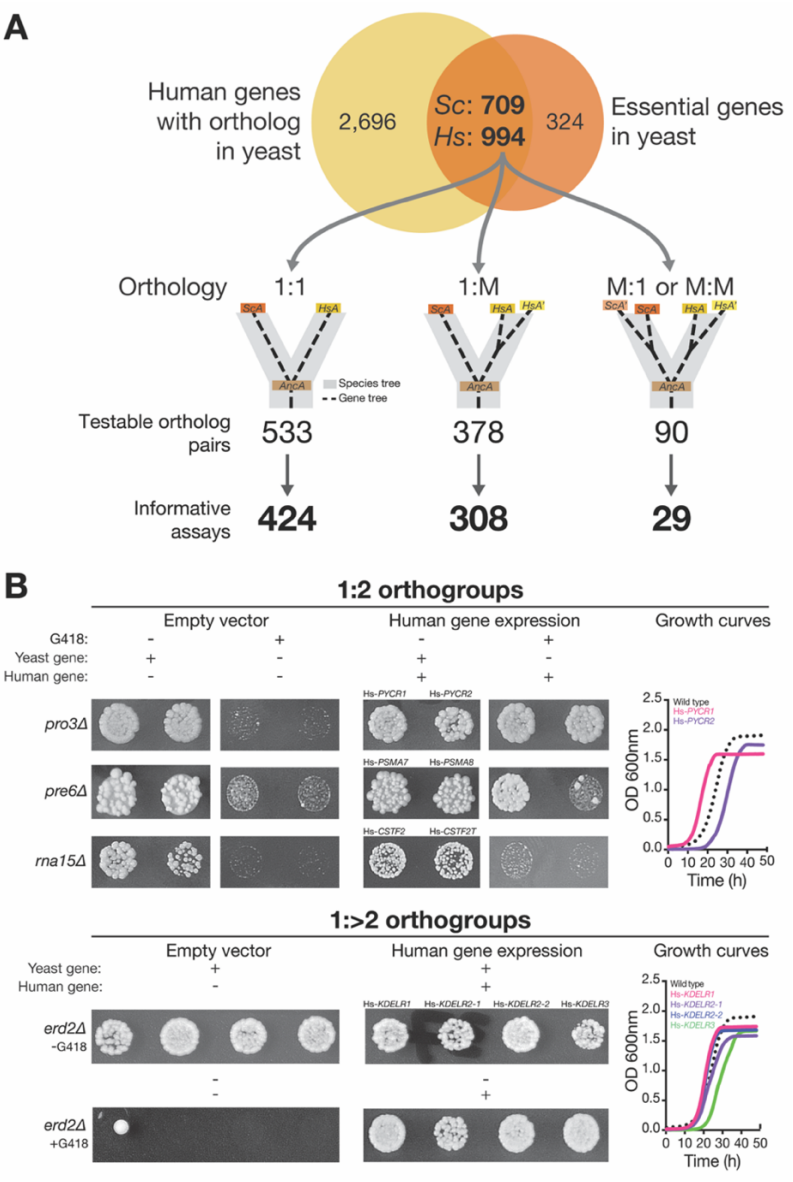


Figure 2.6.1. Systematic functional replacement of essential yeast genes with multiple human co-orthologs.

(A) We identified 994 human genes that are orthologs of 709 essential yeast genes. Of these ortholog pairs, we had previously obtained results for 424 pairs with no

duplications in either yeast or human lineage (i.e., with 1:1 orthology) (Kachroo *et al.* 2015). In this study, we tested the remaining set of essential yeast genes that have acquired lineage-specific duplications, classifying them as 1:M (1 yeast to 2 or more human co-orthologs) or M:M (≥2 yeast to ≥2 human co-orthologs) or M:1 (≥2 yeast to 1 human ortholog). There are 140 essential yeast genes with more than one human ortholog, representing 378 ortholog pairs to be tested. In the case of the 1:M category, we obtained 308 informative assays out of 378 testable pairs, whereas in the case of the M:M or M:1 set, we had 29 informative assays out of 90 testable pairs. Replaceability assays were performed in both the hetKO collection and the temperature-sensitive haploid yeast collection. (B) Representative assays performed in yeast hetKO strains for 1:2 (top) and 1:>2 (bottom) are shown. HetKO yeast strains expressing human genes were sporulated, and the sporulation mix was spotted on Magic Marker medium (see Materials and methods) with (yeast gene absent) or without (yeast gene present) G418. Assays were performed with empty vector control (human gene absent) or yeast expression vectors carrying a human cDNA (human gene expression). In the 1:2 class, three different outcomes of human gene replaceability in yeast were obtained. Top panel: both human co-orthologs (Hs-*PYCR1* and Hs-*PYCR2*) can replace their yeast equivalent (Sc-*PRO3*). Middle panel: one of the two human co-orthologs (Hs-*PSMA7* but not Hs-*PSMA8*) can replace its yeast equivalent (Sc-*PRE6*). Bottom panel: neither of the two human co-orthologs (Hs-*CSFT2* or Hs-*CSFT2T*) can replace its yeast equivalent (Sc-*RNA15*). In the 1:>2 class, an example of all human co-orthologs (Hs-*KDELR1*, two variants of Hs-*KDELR2*, and Hs-*KDELR3*) replacing their yeast gene counterpart equally well is shown. The yeast growth assays for these replaceable human genes are shown on the right for

each. Haploid yeast gene deletion strains carrying plasmids expressing functionally replacing human genes (colored solid lines) generally exhibit comparable growth rates to the wild-type parental yeast strain BY4741 (black dotted lines). Plotted growth curves display the mean of triplicate growth experiments. 1:M, one-to-many; hetKO, heterozygous diploid knockout; M:1, many-to-one; M:M, many-to-many.

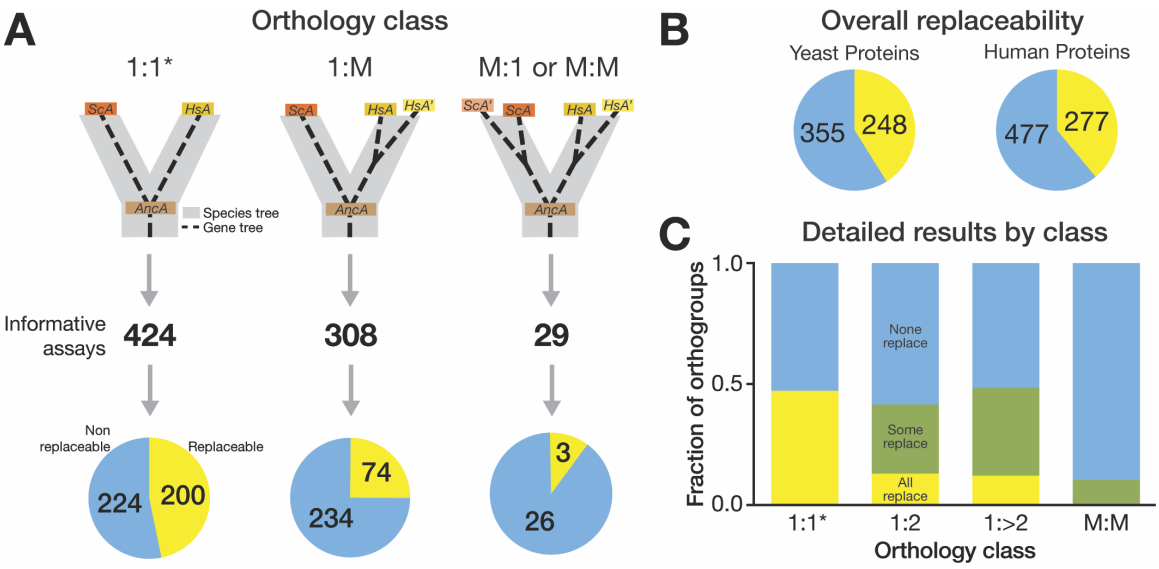


Figure 2.6.2. Distribution of replaceability across orthology classes.

Only rarely did all human co-orthologs in one orthogroup replace. Rather, a family of human proteins typically had one or a few replaceable members or none at all. (A) Previously, systematic replacement of essential yeast genes with 1:1 orthologs (1 yeast to 1 human) demonstrated nearly 50% replaceability of essential yeast genes (Kachroo *et al.* 2015). Here, we tested the replaceability of essential yeast genes with their human counterparts that have acquired lineage-specific duplications in either yeast or human lineage. Of the 308 informative assays obtained in the 1:M class (1 yeast to 2 or more human co-orthologs), 74 human genes replaced their yeast equivalents, whereas 234 did not. Of the 29 informative assays obtained in the M:1 and M:M class (≥2 yeast to 1 or more human co-orthologs), three human genes replaced their yeast equivalents, whereas 26 did not. (B) Combining our previous replaceability assays (Kachroo *et al.* 2015) with the assays done in this study, we have identified 248 essential yeast genes that are functionally replaceable by their human counterparts and 355 that are not. From the

perspective of human proteins, 277 replace their yeast versions, whereas 477 do not. Summary of all the humanized yeast assays performed thus far. (C) Nearly half of the essential yeast genes belonging to the 1:1 orthology class were replaceable by their human equivalents (yellow). The distribution of essential yeast genes replaced by at least one human ortholog in the 1:M (both 1:2 and 1:>2 combined) closely matched the 1:1 results. These yeast genes were rarely replaced by all human co-orthologs in an orthogroup (yellow), with the majority of replaceability falling in the differentially replaceable set (green). M:M yeast orthologs were rarely replaceable by any human ortholog, with only three (of 29 tested) human genes replacing three separate yeast orthologs. Nonreplaceable genes are indicated in blue. * indicates data from previous work (Kachroo *et al.* 2015). 1:M, one-to-many; M:1, many-to-one; M:M, many-to-many.

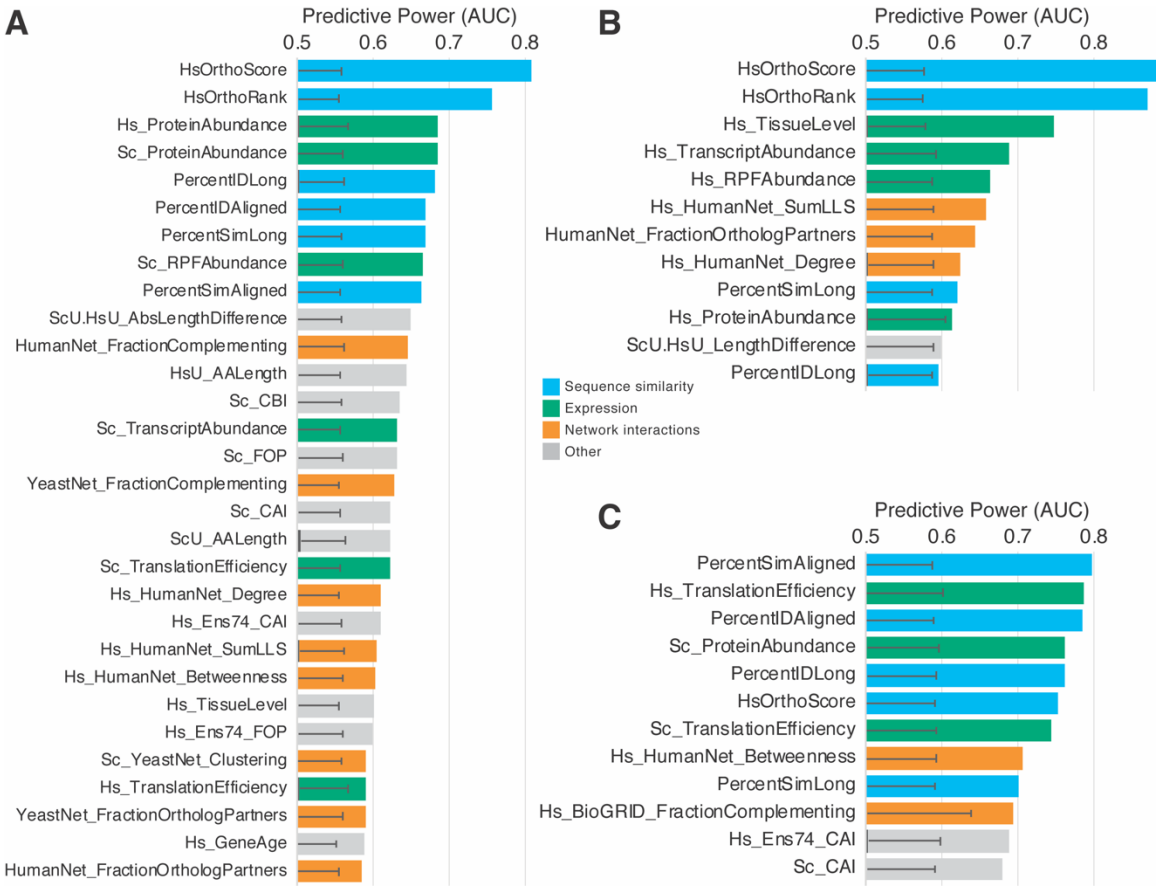


Figure 2.6.3. Replaceability of 1:2 orthologs is explained largely by relative divergence of human co-orthologs.

(A) AUCs for the top 30 predictive features for median-collapsed 1:2 orthogroups are shown. The top two predictive features (HsOrthoScore and HsOrthoRank) for this class indicate that replaceability is driven largely by the nearness of the replacing human co-ortholog to the yeast gene relative to a nonreplacing co-ortholog (i.e., most 1:2 orthogroups have only one replacing human co-ortholog, and it is almost always the least-diverged one). (B) The observed trend is even more strongly demonstrated when analysis is restricted to the specific set of 1:2 orthogroups that displays differential replaceability (i.e., one human co-ortholog replaces, and the other does not), with AUCs

nearing 0.9. The extent of tissue-specific expression also becomes significantly predictive in this set, indicating that human co-orthologs that are more broadly expressed are more likely to replace than their more tissue-specifically expressed co-ortholog. The top 12 features are plotted. (C) More-diverged human co-orthologs in a 1:2 pair do replace in several cases (mostly those in which both co-orthologs replace). When restricting analysis to this set, it is apparent that the most predictive property is sequence similarity to the yeast ortholog, along with proteins that are translated efficiently and are typically in high abundance. Black overlapping bars indicate mean, and error bars indicate standard deviation for 1,000 shuffled AUC calculations for each feature. AUCs were calculated for n = 208 1:2 ortholog pairs, collapsed to 117 and 46 median metaorthologs ([A] and [B], respectively) depending on feature availability. AUCs for MDO assays in (C) were calculated for up to 104 MDO pairs (Materials and methods, **Table 2.7.3**). The top 12 features are plotted. For full source data, see supporting information AUC, area under the receiver–operator characteristic curve.

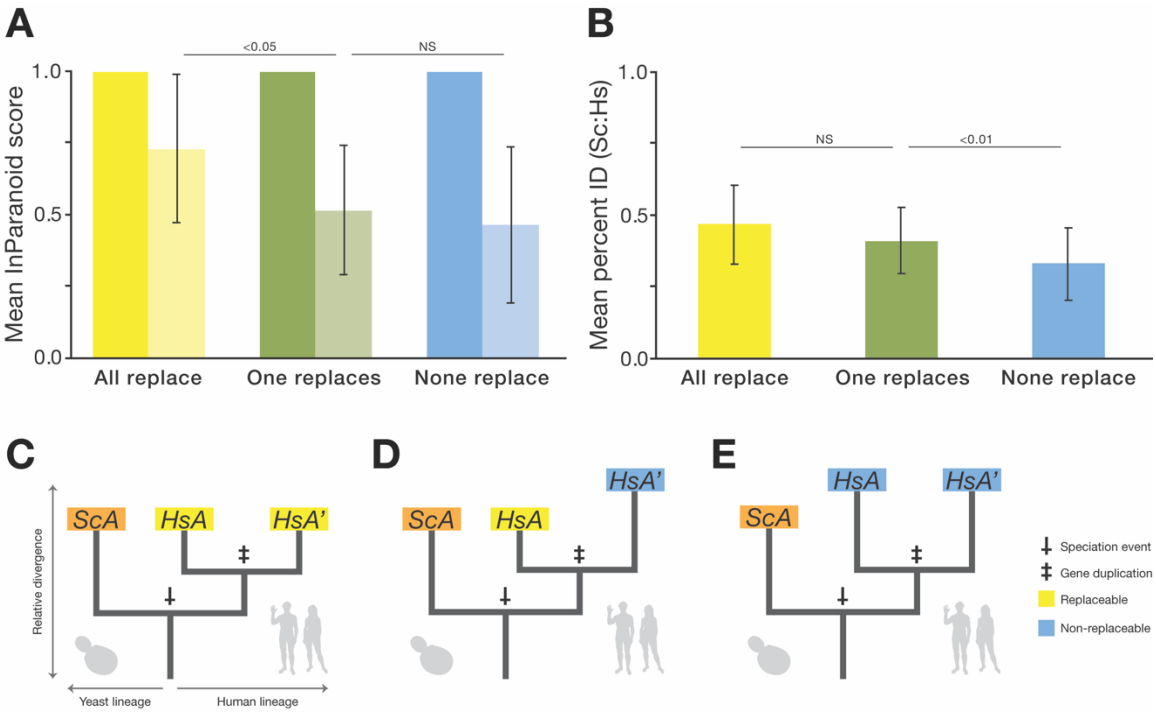


Figure 2.6.4. Replaceability is explained by relative divergence of 1:2 human co-orthologs from each other and their yeast ortholog.

(A) The average ortholog score of the more-diverged (lighter color) 1:2 co-ortholog is more similar to the less-diverged co-ortholog for co-ortholog pairs that both replace than pairs in which only one or neither of the co-orthologs replace. (B) The average percent amino acid identity for 1:2 co-orthologs is not significantly different between the "all replace" (yellow) class and the "one replaces" (green) class, whereas the "none replace" (blue) class is slightly, but significantly, lower than either. Error bars indicate standard deviation. Raw data for (A) and (B) are available with the supporting information. (C–E) Phylogenetic models depicting gene trees for generic 1:2 orthogroups in the various replaceability classes. In (C), the human co-orthologs belong to the "both replace" class and are thus less diverged from each other but are on average similarly diverged from the

yeast ortholog as the “one replaces” class (D), in which the co-ortholog more closely related to the yeast gene is the one that replaces. In (E), the “none-replace” co-orthologs are on average more diverged from the yeast gene. NS, not significant.

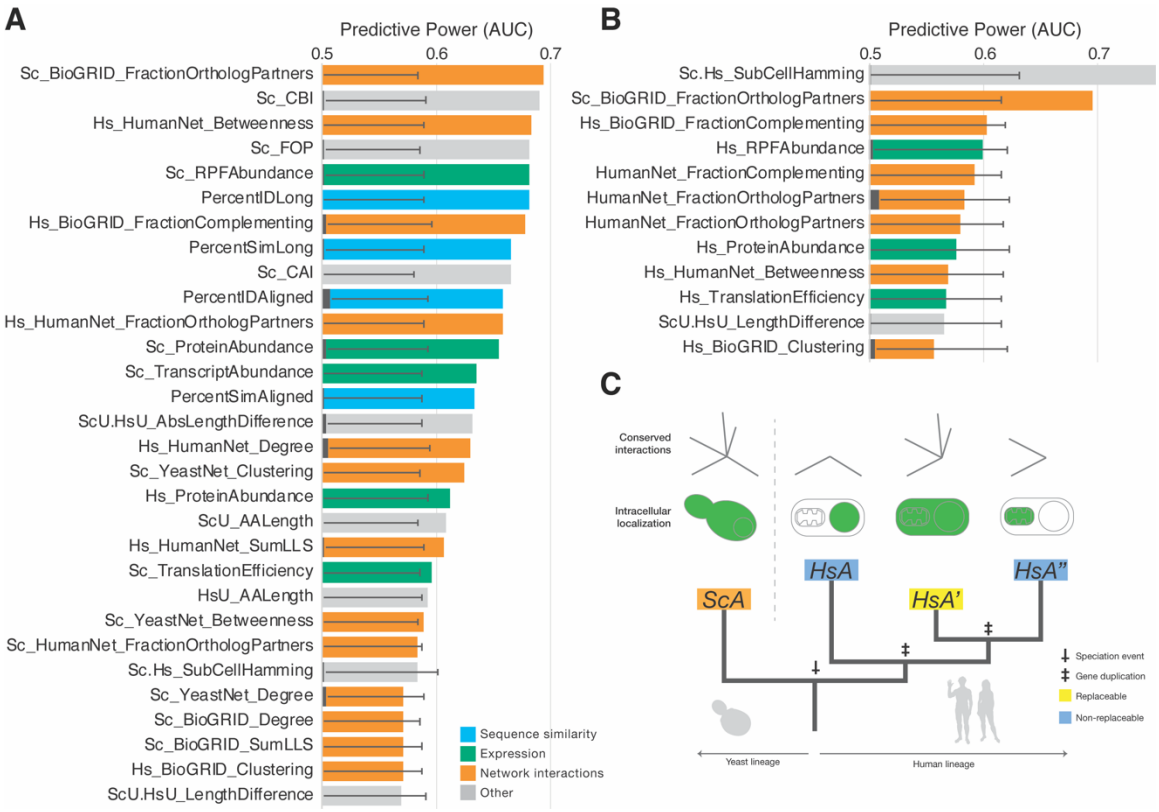


Figure 2.6.5. Replaceable 1:>2 human co-orthologs retain orthologous interaction partners and are more central in interaction networks.

(A) AUCs of the top 30 predictive features for median-collapsed informative 1:>2 co-ortholog pairs. The top two most significant features demonstrate the importance of network context in retaining ancestral functions. Specifically, human co-orthologs in highly expanded gene families that have retained a higher fraction of orthologous protein interaction partners with their yeast ortholog (FractionOrthologPartners) are more likely to replace, as well as those that maintain higher centrality in functional interaction networks (betweenness). (B) Similar to the 1:2 case, we further restricted our analysis to a subset of median-collapsed 1:>2 orthogroups that had both replaceable and nonreplaceable human co-orthologs. In this set, the subcellular localization of the human

proteins appears to be predictive, in that more broadly localized co-orthologs are more likely to replace than their more organellar-specific co-orthologs. Although this AUC is not significant (indicated by being just less than 2 standard deviations off the mean), it is an obvious trend and does overtake fraction of orthologous partners as the highest performing AUC for this set. (C) Phylogenetic model of a generic 1:M orthogroup showing that the replaceable human co-ortholog has retained more orthologous interactions in a network and is localized in a similar manner to the yeast ortholog. For AUC bar plots, black overlapping bars indicate mean, and error bars indicate standard deviation for 1,000 shuffled AUC calculations for each feature. AUCs were calculated for n = 170 1:>2 ortholog pairs, collapsed to 49 and 26 median metaorthologs ([A] and [B], respectively) depending on feature availability (Materials and methods, Table 2.7.3). The top 12 features are plotted in (B). For full source data, see supporting information supporting information. 1:M, one-to-many; AUC, area under the receiver–operator characteristic curve.

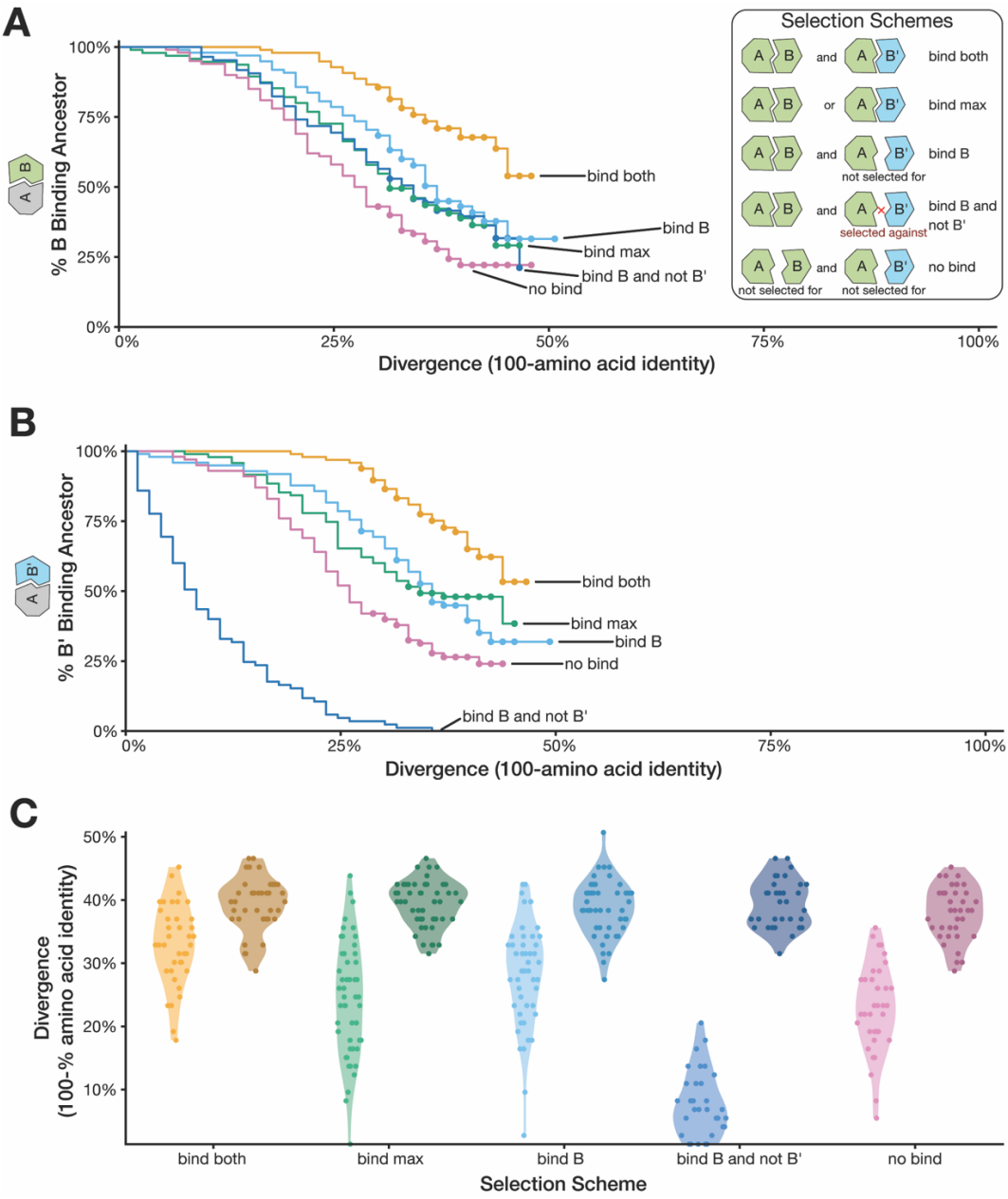


Figure 2.6.6. Simulated protein evolution suggests that diverged duplicates are less likely to bind their ancestral interaction partner.

(A) Inset: Five different types of selection scenarios were considered for a heterodimeric protein complex, considering the effects of amino acid substitutions using the Rosetta

molecular modeling platform. (1) AB and AB′ (bind both). (2) AB or AB′, and only the most stable binding interface is considered (bind max). (3) AB, but AB′ was not enforced (bind B). (4) AB, and AB′ is selected against (bind B and not B′). (5) Neither AB nor AB′ is selected for (no bind). Percent of simulations in which an evolved B subunit has the ability to bind the ancestor to A. Divergence is measured by the amount B has diverged from the ancestor of B. (B) Percent of simulations in which an evolved B′ subunit has the ability to bind the ancestor to A. Divergence is measured by the amount B′ has diverged from the ancestor of B. (C) Percent divergence of duplicates when only B or B′ is able to bind the ancestor A. Lighter hues denote a duplicate that is able to bind the ancestor of A, and darker hues denote the nonbinding duplicate. Data and scripts for these are available at the following link:https://github.com/a-teufel/Laurent_etal_2020. Figure A.8.dapted from(Teufel *et al.* 2019).

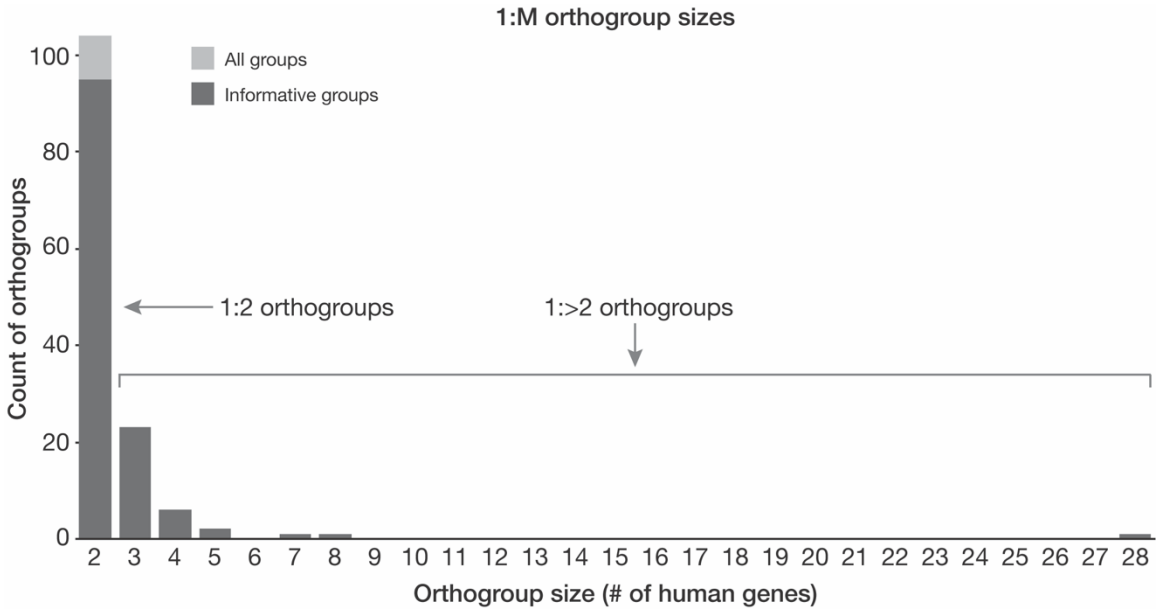
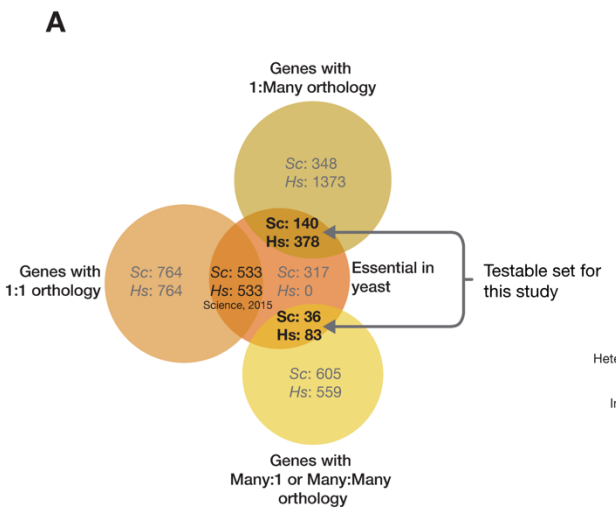


Figure 2.6.S1. Count of orthogroups with the corresponding number of human gene members.

The vast majority of one to multiple orthologs are of the 1:2 variety (1 yeast gene to 2 human co-orthologs), whereas the rest were classified as 1:>2 (1 yeast gene to >2 human co-orthologs). These two groups were considered separately during feature analysis.



B

	1:M (Sc:Hs)		M:M (Sc:Hs)
	1:2	1:>2	
Human ORFeome clones available	182	156	43
Human ORFeome sequence verified	170	141	33
MGC sequence verified human cDNA	75	57	7
Total testable human expression clones	245	198	40
Replaceability assays in Heterozygous diploid knockout strains	220	189	40
Informative assays in Heterozygous diploid knockout strains	214	161	29
Assays in TS strains	131	104	0
Informative assays in TS strains	98	92	0

Figure 2.6.S2. Detailed overview of the ortholog groups in this study.

(A) Venn diagram indicating the overlap of essential yeast genes with their human orthologs in various ortholog classes. Ortholog definitions are based on InParanoid (Remm *et al.* 2001), whereas essentiality of yeast genes is based on the systematic yeast knockout study by Winzeler and colleagues (Winzeler *et al.* 1999). (B) A thorough breakdown of the numbers of ortholog pairs identified through to those resulting in informative assays.

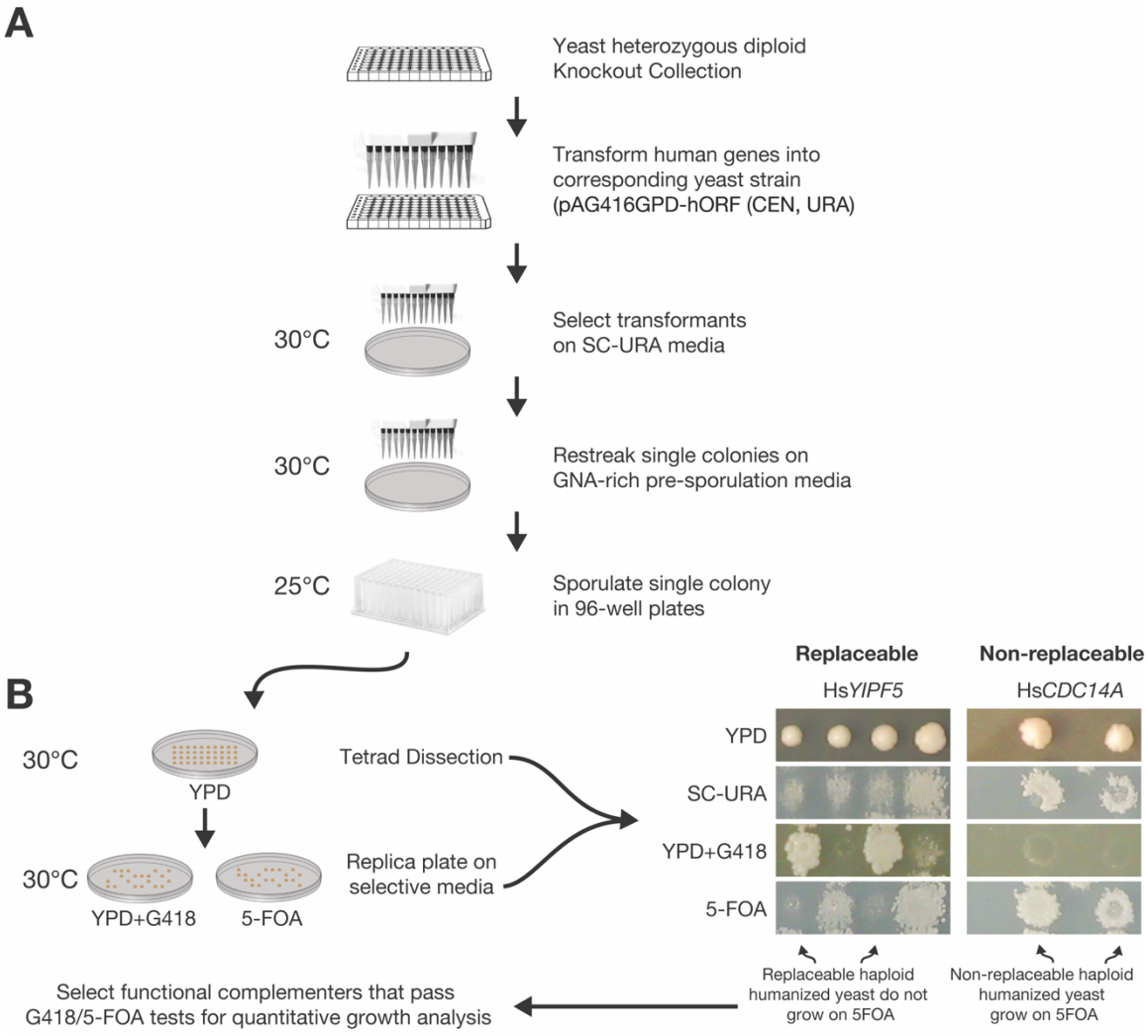


Figure 2.6.S3 Detailed illustration of yeast gene replaceability, tetrad dissection, and plasmid loss assays performed in a yeast hetKO collection.

(A) Yeast strains were grown in 96-well plates in selective medium (YPD + 200 μg/ml G418). The matched orthologous human gene expression clones in 96-well plates were transformed into the appropriate yeast strains, followed by selection on the appropriate dropout medium (SD-URA). The resulting transformants were spotted onto selective medium with appropriate markers to assay complementation. (B) In the case of the yeast

hetKO collection, replacement was further verified by carrying out tetrad dissection followed by testing for plasmid dependency (selection on 5-FOA). Representative tetrad dissection assays are shown (right). Hs-YIPF5 functionally replaces its yeast counterpart. Dissected tetrads showed 2:2 segregation, with tetrads 1 and 3 containing the wild-type yeast allele (surviving on 5-FOA), whereas tetrads 2 and 4 contain the KanMX allele (yeast null allele) complemented by the human gene (surviving on YPD + G418). We used 5-FOA was for counterselecting the human gene encoding plasmid, whereas YPD + G418 was used to select for the yeast null allele (KanMX selection), respectively. However, in cases on noncomplementers (e.g., Hs-*CDC14A*), only two yeast spores (carrying the wild-type yeast allele) survive on YPD and fail to grow on YPD + G418 but grow on 5-FOA, indicating plasmid loss. (C) Each confirmed humanized haploid yeast strain was assayed for growth defects to quantify the replaceability. hetKO, heterozygous diploid knockout.

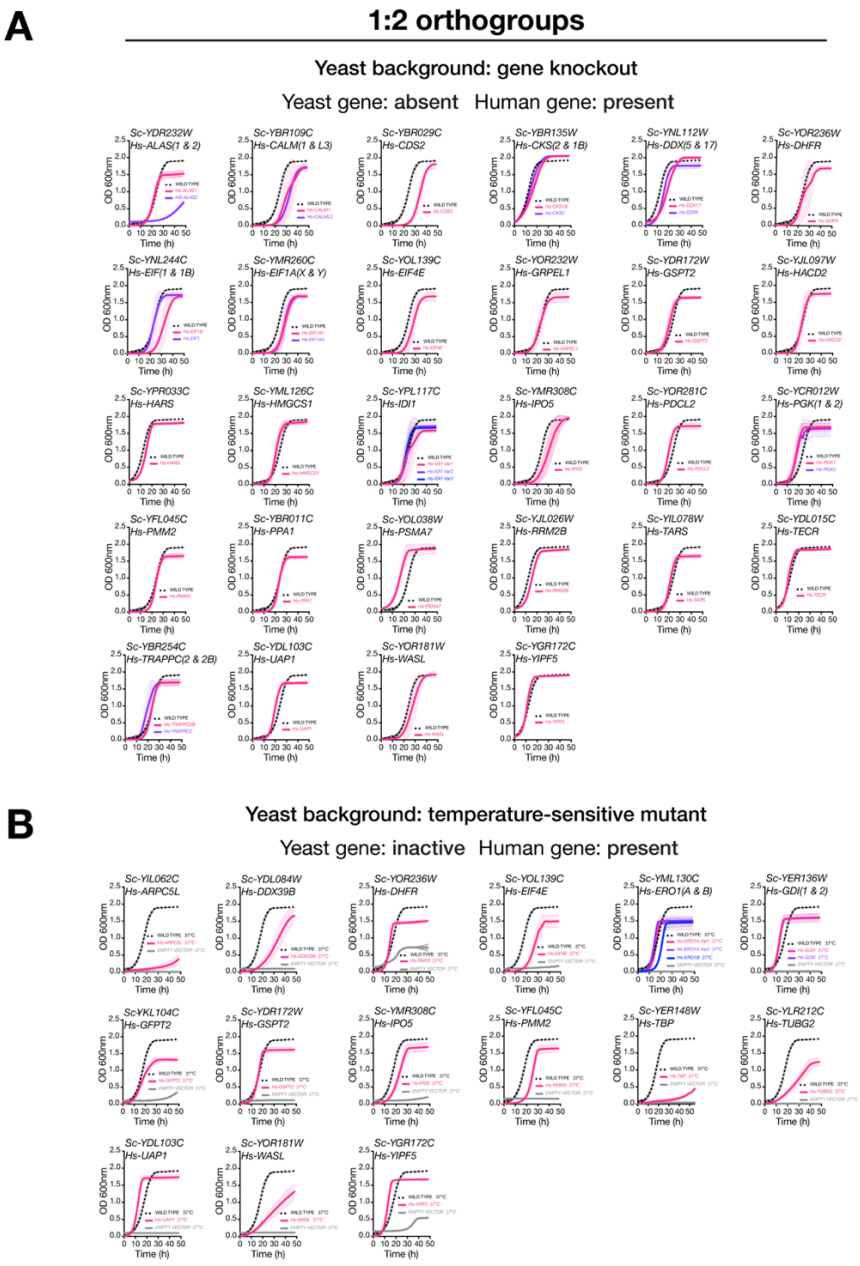


Figure 2.6.S4. Quantitative growth assays of all humanized yeast strains belonging to the 1:2 class.

(A) Growth curves of humanized haploid yeast strains obtained after sporulation and tetrad dissection and followed by 5-FOA screening of yeast hetKO strains. The assays

were performed in the presence of G418 (yeast gene–absent and human gene–present condition). Haploid yeast gene deletion strains carrying plasmids expressing functionally replacing human genes (colored red or purple solid lines) generally exhibit comparable growth rates to the wild-type parental yeast strain BY4741 (black dotted lines), except in the case of Hs-*ALAS2*, which shows reduced ability to replace the orthologous yeast gene compared with Hs-*ALAS1*. Mean and standard deviation plotted from triplicate assays. (B) Growth curves of humanized haploid temperature-sensitive yeast strains performed at a restrictive temperature of 37 °C (yeast gene–inactive and human gene–present condition). Temperature-sensitive haploid yeast strains carrying plasmids expressing functionally replacing human genes (colored red, purple, or blue solid lines) generally exhibit comparable growth rates to the wild-type parental yeast strain BY4741 (black dotted lines). The yeast strains harboring empty vector without a corresponding human gene, showing no or poor growth at restrictive temperatures of 37 °C (gray solid lines), serve as controls. In total, 208 individual assays were performed (as two independent biological replicates). In all, 45 human genes showed functional replaceability in either assay, whereas 127 did not, and 36 human gene complementation assays were noninformative (cases in which the control experiments did not behave appropriately). Raw data for growth curves are available with the supporting information. hetKO, heterozygous diploid knockout.

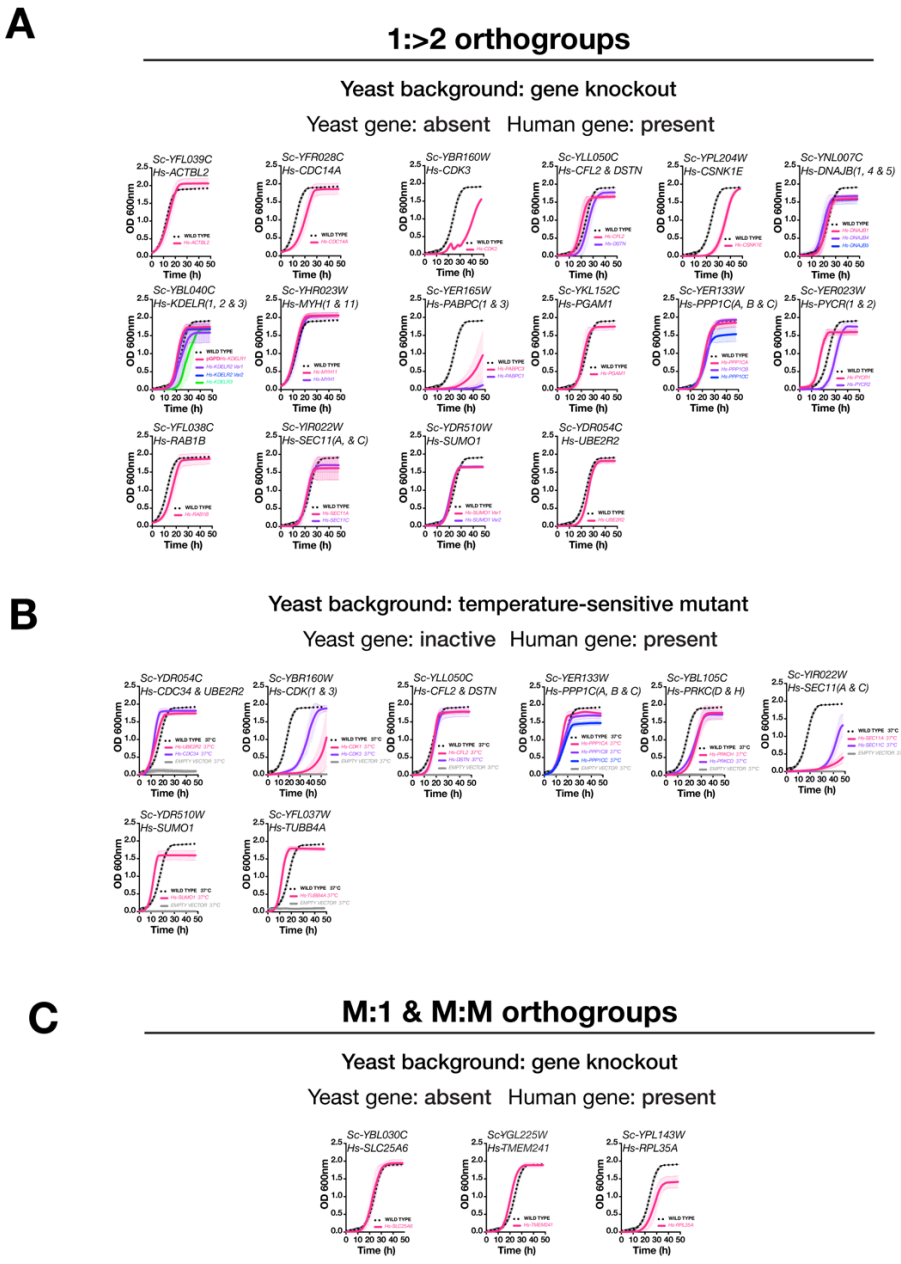


Figure 2.6.S5. Quantitative growth assays of all humanized yeast strains belonging to 1:>2, M:M, and M:1 orthogroups.

In total, 170 individual assays were performed (as two independent biological replicates):

32 human genes showed functional replaceability in either assay, whereas 105 did not,

and 33 human gene replaceability assays were noninformative (cases in which the control experiments did not behave appropriately). (C) Growth curves of humanized haploid yeast strains (M:1 and M:M class) obtained after sporulation and/or tetrad dissection and followed by 5-FOA screening of hetKO strains. The assays were performed in the presence of G418 (yeast gene–absent and human gene–present condition). In total, 40 individual assays were performed (as two independent biological replicates). Three human genes showed functional replaceability, whereas 26 did not, and 11 human gene replaceability assays were noninformative (cases in which the control experiments did not behave appropriately). Mean and standard deviation plotted with $N = 3$. (A) Growth curves of humanized haploid yeast strains (1:>2 class) obtained after sporulation and/or tetrad dissection followed by 5-FOA screening of hetKO strains. The assays were performed in the presence of G418 (yeast gene–absent and human gene–present condition). Haploid yeast gene deletion strains carrying plasmids expressing functionally replacing human genes (colored red, purple, and blue solid lines) generally exhibit comparable growth rates to the wild-type parental yeast strain BY4741 (black dotted lines), except in the case of Hs-*PABPC1* and Hs-*PABPC3*, which shows reduced ability to replace the orthologous yeast gene. Mean and standard deviation plotted from triplicate assays. (B) Growth curves of humanized haploid temperature-sensitive yeast strains (1:M class) performed at a restrictive temperature of 37 °C (yeast gene–inactive and human gene–present condition). Temperature-sensitive haploid yeast strains carrying plasmids expressing functionally replaced human genes (colored red, purple, or blue solid lines) generally exhibit growth rates comparable to their parental wild-type yeast strain BY4741 (black dotted lines), except in the cases of Hs-*CDK1*, Hs-*SEC11A*, and Hs-*SEC11C*,

which showed reduced ability to functionally replace their orthologous yeast gene. Yeast strains harboring an empty vector without a corresponding human gene, showing no or poor growth at a restrictive temperature of 37 °C (gray solid lines), serve as controls. (Raw data for growth curves are available in the supporting information) 1:M, one-to-many; hetKO, heterozygous diploid knockout; M:1, many-to-one; M:M, many-to-many.

A

OrthoGroup	CompStatus	HsOrthoScore	HsOrthoRank	PercentIDLong	PercentSimLong	ScU_AALength	HsU_AALength
218	Non complement	1	1	0.51476015	0.78413284	542	536
218	Non complement	0.714	2	0.51660517	0.78782288	542	536
218	Non complement	0.644	3	0.50738007	0.78413284	542	538
249	Complement	0.482	2	0.44025157	0.66352201	577	636
249	Complement	0.405	4	0.42789223	0.65134707	577	631
249	NA	1	1	0.43485342	0.67589577	577	614
249	Non complement	0.408	3	0.43944099	0.66304348	577	644

B

OrthoGroup	CompStatus	HsOrthoScore	HsOrthoRank	PercentIDLong	PercentSimLong	ScU_AALength	HsU_AALength
218	Non complement	0.714	2	0.51476015	0.78413284	542	536
249	Complement	0.4435	3	0.4340719	0.65743454	577	633.5
249	Non complement	0.408	3	0.43944099	0.66304348	577	644

Figure. 2.6.S6. Demonstration of the "median-collapse" feature table procedure.

(A) An example subset of full ortholog data for two orthogroups in the 1:M ortholog class. (B) The same data as in (A) following median collapse. Orthogroup 218 has been collapsed to a single metaortholog pair with the status "noncomplement" because all human co-orthologs failed to complement. Group 249 has been collapsed to two metaortholog pairs because there were co-orthologs that complemented and did not. Metaortholog feature values are the medians of the full ortholog values with the same status. 1:M, one-to-many

2.7. TABLES

Owing to their extensive size and large format, all tables and supporting information are provided with the supplementary materials. Additionally, they can also be in this website.

Table 2.7.1. Complementation results for all assays.

We list the results of each assay background (TS or MM) as well as tetrad complementation reassay status and the final overall status used for subsequent analyses. Also listed is the literature status of genes found in YeastMine as previously reported literature assays. Those without a literature status are considered novel to this study. TS, temperature-sensitive.

Table 2.7.2. Literature assay status from YeastMine.

We list the results of 1:M ortholog assays previously reported in the literature as found via the YeastMine database. We also indicate whether these match our reported assays and the accuracy calculation reported in the text. 1:M, one-to-many.

Table 2.7.3. Protein and ortholog properties.

Includes all properties assembled and calculated for all proteins or ortholog pairs used for ROC analyses. ROC, receiver–operator characteristic.

Chapter 3: Systematic humanization of the yeast cytoskeleton discerns functionally replaceable from divergent human genes[2]

One of the central themes of my graduate work here has been understanding the principles that underlie cross-species gene complementarity and how genes diverge over evolution. Until recently, the majority of the humanization assays in yeast surveyed the ability of human gene families to complement a lethal defect induced by the loss of their corresponding yeast ortholog. However, assaying conditionally essential roles and/or context-dependent non-essential roles were only being studied to a limited extent in the lab. This prompted us to study this question of complementarity beyond growth assays to more fine-grained phenotypes.

This chapter explicitly focuses on the core structural elements of the yeast cytoskeleton including the actin, myosin, tubulin, and septin gene families, testing the abilities of human genes to complement defects in a rich assortment of critical cellular roles caused by loss of their yeast counterparts. We observe a striking pattern of replaceability across different ortholog groups, with differential replaceability being a noticeable trend, seldom observing replaceability by all human members within the family. In addition to the rescue of growth defects, we also probe functional complementation in more detail. We observe varying extents of the functional rescue of the deletion of yeast genes by their human orthologs corresponding to cytoskeletal

[2]This chapter was published as Garge, R.K., Laurent, J.M., Kachroo, A.H., and Marcotte, E.M. (2020). "Systematic Humanization of the Yeast Cytoskeleton Discerns Functionally Replaceable from Divergent Human Genes." And was featured on the cover of *Genetics* 215, 1153–1169. My contributions were carrying out the complementation experiments, analysis and writing of the manuscript while being supported by Jon M. Laurent and Aashiq H. Kachroo.

family-specific roles, specifically cell morphology, meiosis, and mating. By comparing distinct morphological phenotypes observed in humanized strains with their respective yeast interaction network(s), we identify replaceable human orthologs capable of executing essential yeast roles but seemingly breaking key non-essential interactions crucial for maintaining cell morphology.

3.1. ABSTRACT

Many gene families have been expanded by gene duplications along the human lineage, relative to ancestral opisthokonts, but the extent to which the duplicated genes function similarly is understudied. Here, we focused on structural cytoskeletal genes involved in critical cellular processes including chromosome segregation, macromolecular transport, and cell shape maintenance. To determine functional redundancy and divergence of duplicated human genes, we systematically humanized the yeast actin, myosin, tubulin, and septin genes, testing ~81% of human cytoskeletal genes across 7 gene families for their ability to complement a growth defect induced by inactivation or deletion of the corresponding yeast ortholog. In 5 of 7 families—all but α-tubulin and light myosin, we found at least one human gene capable of complementing loss of the yeast gene. Despite rescuing growth defects, we observed differential abilities of human genes to rescue cell morphology, meiosis, and mating defects. By comparing phenotypes of humanized strains with deletion phenotypes of their interaction partners, we identify instances of human genes in the actin and septin families capable of carrying out essential functions, but apparently failing to fully complement the cytoskeletal roles

of their yeast orthologs, thus leading to abnormal cell morphologies. Overall, we show that duplicated human cytoskeletal genes appear to have diverged such that only a few human genes within each family are capable of replacing the essential roles of their yeast orthologs. The resulting yeast strains with humanized cytoskeletal components now provide surrogate platforms to characterize human genes in simplified eukaryotic contexts.

3.2. INTRODUCTION

Gene duplication is regarded as one of the key drivers of evolution, contributing to the generation and accumulation of new genetic material within species (Koonin 2005; Soukup). Duplication creates an initial multiplication of dosage and functional redundancy, but the trends dictating how duplicated genes retain function or diverge are still unclear. The processes governing the distribution of molecular roles within gene families also have important consequences for annotating genes, which generally takes advantage of sequence similarity and conservation over vast timescales of divergence to infer functions of homologous genes across species. Many individual studies have directly tested the conservation of function among orthologs from different species by swapping them from one species into another (Heinicke *et al.* 2007; Cherry *et al.* 2012). However, only recently have efforts been made to test such functional equivalence more systematically, with several recent large-scale studies harnessing "the awesome power of yeast genetics" to systematically replace yeast genes by their human, plant, or even bacterial counterparts and assay for functional compatibility (Kachroo *et al.* 2015, 2017;

Hamza *et al.* 2015; Laurent *et al.* 2016; Sun *et al.* 2016; Yang *et al.* 2017). Although humans and yeast last shared a common ancestor nearly a billion years ago, these studies have demonstrated that substantial fractions (12-47%) of tested essential yeast genes could be replaced by their human equivalents (Kachroo *et al.* 2015, 2017; Hamza *et al.* 2015; Sun *et al.* 2016; Yang *et al.* 2017; Laurent *et al.* 2019). The ability of many human genes to functionally replace their yeast orthologs demonstrates the high degree of functional conservation in eukaryotic systems over billion year evolutionary timescales (Kachroo *et al.* 2015; Laurent *et al.* 2016).

Previous humanization efforts in yeast have primarily focused on ortholog pairs with no obvious duplications within yeast and human lineages (1:1 orthologs), only partially testing the orthologs in expanded gene families (Kachroo *et al.* 2015; Hamza *et al.* 2015; Sun *et al.* 2016; Yang *et al.* 2017; Laurent *et al.* 2019) and seldom beyond assaying impact on growth rate. In this study, to better understand functional conservation across expanded gene families in core eukaryotic processes, we focused on the major structural components of the eukaryotic cytoskeleton, including actins, myosins, septins, and tubulin genes. Genes constituting the eukaryotic cytoskeleton play key roles in critical cellular processes, mainly organizing the contents of the cell by dynamically controlling cell shape, positioning organelles, and transporting macromolecules including chromosomes across the cell through the generation of coordinated mechanical forces (Wickstead and Gull 2007, 2011; Wickstead *et al.* 2010). Importantly, cytoskeletal gene families have undergone large expansions along the human lineage, while being restricted to only a few family members in yeast (**Fig. 3.6.1**).

Advances in comparative genomics have shed light on the likely cytoskeletal components present in the last eukaryotic common ancestor (LECA)(Wickstead *et al.* 2010; Wickstead and Gull 2011). Additionally, phylogenetic profiling studies of the cytoskeleton have been highly informative in inferring gene loss and retention events across eukaryotic clades (Wickstead *et al.* 2010; Wickstead and Gull) (**Fig. 3.6.1**). Such studies suggest that the origins of the eukaryotic cytoskeleton predate eukaryogenesis and ancestrally trace back to primitive tubulin- and actin-like homologs in bacteria[13]. These components critical in cell division subsequently evolved to incorporate families of accessory motors and regulatory proteins expanding towards performing vital cellular roles, including phagocytosis, motility and vesicular transport, still evident across vast eukaryotic clades of life (McKean *et al.* 2001; Wickstead *et al.* 2010; Janke 2014; Jékely 2014; Hall and Russell).

Though the cellular roles of human cytoskeletal gene families have been broadly elucidated, aided by studies in simpler eukaryotes (Schatz *et al.* 1986a; Winsor and Schiebel 1997; Bode *et al.* 2003; Nogales *et al.* 2010; Li *et al.* 2011), the specific functions of their constituent family members in humans have to date still only been partially characterized (Tischfield *et al.* 2010; Feng *et al.* 2016; Stottmann *et al.* 2016; Wawro *et al.* 2017; Huang *et al.* 2018b; Yuan *et al.* 2018; Chen *et al.* 2019). Functional assays in human cell lines pose the challenge of functional redundancy, with buffering by other paralogs complicating the determination of paralog-specific roles within cytoskeletal gene families. The high degree of sequence conservation among paralogs within each cytoskeletal family make functional analysis of individual cytoskeletal genes directly in human cells both experimentally and computationally cumbersome. However,

cross-species gene swaps have the potential to provide direct assays of individual paralogs within these expanded gene families, thereby revealing the extent to which present day orthologous genes retain ancestral function.

To understand the extent to which human cytoskeletal genes in expanded orthogroups retain cross-species functional equivalence, we systematically humanized major elements of the yeast cytoskeleton. We tested ~81% (50/62) of all human genes from actin, myosin, septin, and tubulin families, using a combination of classical yeast genetics and CRISPR-Cas9 mediated genome editing to assay the replaceability of essential cytoskeletal orthologs, as initially determined *via* simple growth rescue complementation assays. Overall, we show that (13/50) members from 5 of 7 tested gene families (actin, heavy myosin, septin, β- and ɣ-tubulin) can indeed execute essential roles of their yeast counterparts. Within each replaceable family we show that several present-day human orthologs still possess functional roles of their respective opisthokont ancestors compatible in a yeast cellular context. Additionally, we characterized cellular phenotypes beyond growth and observed differential abilities among complementing human cytoskeletal genes to carry out non-essential cytoskeletal roles, including those in cell morphology, sporulation, mating, meiosis, and cytokinesis. Besides revealing human cytoskeletal orthologs capable of performing their core eukaryotic roles, yeast strains with human cytoskeleton components additionally serve as cellular reagents to study the specific roles of expanded cytoskeletal gene family members in a simplified unicellular eukaryotic context.

3.3. MATERIALS AND METHODS

3.3.1. Curating human orthologs to yeast genes

Yeast genes to be tested for replaceability were curated from Saccharomyces Genome Database (Cherry *et al.* 1998, 2012), and human orthologs were curated from the InParanoid (Sonnhammer and Östlund 2015) and EggNOG (Huerta-Cepas *et al.* 2016) databases both of which employ graph-based algorithms that recognize orthogroups between species by exhaustively performing an all-vs-all bidirectional BLAST search of all protein sequences and Hidden Markov Models to estimate ortholog groups respectively. We curated a total of 62 human cytoskeletal proteins with yeast orthologs in the 7 major cytoskeletal gene families.

3.3.2. Cloning human cytoskeletal ORFs

Human genes were extracted from the ORFeome (Lamesch *et al.* 2007), a collection of *E. coli* strains, each containing a single human ORF cDNA cloned into a Gateway 'entry' vector (Alberti *et al.* 2007). Human genes cloned in this manner are flanked by attL recombination sites. To generate yeast expression vectors, the human entry vectors were isolated and subjected to Gateway LR reactions with a yeast Gateway 'destination' vectors followed by transformation into competent *E. coli* to obtain yeast expression clones. We used destination vectors from the Advanced Yeast Gateway kit (Alberti *et al.* 2007), specifically pAG416-GPD-ccdB, pAG416-GPD-ccdB-eGFP, pAG426-GPD-ccdB-eGFP, pAG426-GPD-ccdB (CEN, Ura) destination vectors. Since the original version of pAG416-GPD-ccdB does not encode a stop codon immediately after the cloning region (which is also not encoded in ORFeome genes), this results in a

~60 amino acid tail being translated to any protein expressed from it. To eliminate potential issues from the tail, we mutagenized the vector downstream of the cloning region to introduce a stop codon, thereby shortening the tail to six amino acids (pAG416-GPD-ccdB+6Stop) (Kachroo *et al.* 2015). Prior to performing complementation assays, all expression clones were verified by Sanger sequencing to ensure there were no sequence errors or mutations in the human cytoskeletal genes prior to complementation assays in yeast.

3.3.3. Assaying human cross-species complementation in yeast

1. Tetrad dissection and analysis

In all, 40 human genes were assayed using tetrad dissection (**Fig. 3.6.S2A**): The yeast heterozygous diploid deletion strain (Magic Marker) collection (Giaever *et al.* 2002) (obtained from ATCC) with one allele replaced with a Kanamycin-resistance (KanMX) cassette was used. Human genes in Gateway entry clones (pDON223) were curated from the Human ORFeome collection(Lamesch *et al.* 2007) and cloned into yeast destination vectors. We transformed each human clone or an empty vector (pAG416-GPD-ccdB+6Stop lacking the human ORF) control into the appropriate heterozygous diploid deletion strain yeast strain and grew them on SC–Ura + G418 (200µg/ml) to select for the human clone (Ura) and KanMX (G418) simultaneously. Transformants were then plated on GNA-rich pre-sporulation media containing G418 (200µg/ml). Individual colonies were inoculated into a liquid sporulation medium containing 0.1% potassium acetate, 0.005% zinc acetate, and were incubated at vigorous shaking at 25˚C for 3-5 days. Following this, sporulation efficiency was estimated by microscopy, and successful sporulations were subjected to tetrad dissection and analysis using standard

protocols (Dunham *et al.* 2015). We initially plate the spores from tetrads on YPD media. This allows us to check for the spore viability. The colonies are then replica-plated onto either YPD+G418 media or -Ura (+ or - G418) media or 5-FOA plates. We predominantly observe 2 spores with Ura selection that tend to be G418 resistant as well. This behavior could be due to various patterns of segregation of CEN vectors or that the YPD selection allows the Ura to be lost from the wild-type spores. Successful dissections were replica plated both on 5-FOA (for plasmid counter-selection) and YPD + G418 (for yeast null allele selection). Successful complementations consisted of 2:2 segregation with survival on YPD+G418 and failure to grow on 5-FOA (**Fig. 3.6.S2B**). We subsequently performed quantitative growth assays (in triplicate) on the tetrads passing the 5-FOA and G418 segregation test (**Fig. 3.6.S2B, S2C**). Each profile (**Fig. 3.6.S2C**) was analyzed and quantified to detect any growth defects in yeast. 10 out of 40 human genes assayed in this manner functionally rescued yeast from the lethal phenotype. For population-based plasmid complementation assays, we selected the spores on SC-Arg-His-Leu+Can-Ura +/- G418. A complementation consists of spores being able to proliferate on both in the presence (+G418, selecting for spores with the null allele) and absence (-G418, selecting for wild-type spores). Assays where this threshold was not met was deemed non-complementing.

2. Yeast temperature-sensitive (ts) assays

Where available, temperature-sensitive (ts) strains from Li *et al.* (Li *et al.* 2011) were used to assay human gene complementation. These strains grow ideally at lower permissive (22-26˚C) but not restrictive temperatures (35-37˚C). We transformed each

human gene expression vectors and empty vector control plasmids (pAG416-GPD-ccdB+6Stop) into the corresponding ts strains (**Fig. 3.6.S3A**). Transformants were plated on SC-Ura media at both permissive and restrictive temperatures allowing us to control for transformation efficiency and expression toxicity while simultaneously testing for functional replacement by the human gene (**Fig. 3.6.S3B**). Successful complementations consisted of transformants with the human gene surviving at both conditions (growth at permissive or restrictive temperatures) while those with the empty vector failing to survive at restrictive temperatures. Successfully humanized strains were then subjected to quantitative liquid growth assays to monitor robustness of complementation (**Fig. 3.6.S3C**).

3. Genomic replacement via CRISPR-Cas9

29 human genes were assayed using a CRISPR-Cas9 mediated yeast genome editing protocol as previously described (Kachroo *et al.* 2017; Akhmetov *et al.* 2018a). For every yeast gene to be replaced endogenously, a minimum of 2 synthetic guide RNAs (sgRNA) with high on-target and high off-target scores were designed using the Geneious v10.2.6 CRISPR-Cas9 tools suite. Selected sgRNAs were ordered as oligos from IDT and cloned into yeast CRISPR-Cas9 knockout plasmids from the yeast toolkit (YTK) (Lee *et al.* 2015) to express a synthetic guide RNA, Cas9 nuclease and a selectable marker (Ura). Wild-type yeast strains (s288c/BY4741) were transformed with a knockout plasmid and a repair template in the form of a PCR amplicon composed of the human open reading frame flanked with sequence homology to the targeted yeast locus (**Fig. 3.6.S4A)**. Transformants were selected on SC-Ura medium to select for the knockout plasmid.

While CRISPR-mediated double-stranded breaks are intrinsically lethal, targeted editing of an essential gene locus acts as an added layer of selection in replacing the human gene of interest, allowing cells to survive only if the human ortholog being assayed functionally replaces the yeast gene at the appropriate locus (Kachroo *et al.* 2017; Akhmetov *et al.* 2018a). Surviving colonies (**Fig. 3.6.S4B**) obtained in the presence of repair templates were screened for successful humanization by colony PCR using primers outside the region of homology. Confirmed clones were Sanger sequenced and subsequently subjected to quantitative liquid growth assays to evaluate overall fitness. In the case of tubulins, all CRISPR-Cas9 assays were carried out in a BY4741 *tub3Δ* strain to avoid homologous repair of the *TUB1* locus by the *TUB3* gene. To generate diploid strains homozygous for human β-tubulin alleles, we replaced the yeast *TUB2* allele in our heterozygous diploids by re-transforming the CRISPR-Cas9 and sgRNA expression vector specifically targeting yeast β-tubulin allele (*TUB2*) and selected for viable strains without supplying an external repair template, forcing homology -directed repair of the *TUB2* lesion by the human ortholog(s) on the other homologous chromosome.

3.3.4. Growth Assays

Liquid growth assays were performed in triplicate using a Biotek Synergy HT incubating spectrophotometer in 96-well format. All humanized strains were pre-cultured to saturation in YPD and diluted into 150µL of medium to finally have 0.05-0.1 x 10^7 cells/ml. Each growth assay lasted for 48 hrs, with absorbance measured at 600nm every 15 min. For human genes assayed using heterozygous diploid deletion collections, growth assays were performed in YPD, SC-Ura, and YPD+G418 medium to confirm retention of the plasmid and selection of the deletion allele respectively. For CRISPR

assays, humanized strains were grown in YPD. Human genes assayed *via* temperature-sensitive alleles were grown at both temperatures (permissive and restrictive). Growth curve data was processed in Rstudio and plotted using the ggplot2 package or GraphPad Prism.

3.3.5 Microscopy and image analysis

For imaging, for plasmid-based complementation assays, all the yeast strains were grown in selection conditions to maintain the plasmid (Ura+). For the genomically integrated humanized yeast strains, the cells were grown in YPD medium.

For DAPI staining, cells were grown in liquid medium to saturation and fixed in a final concentration of 3.7% (v/v) formaldehyde for 10 minutes at room temperature. Cells were then washed with water and 1M sorbitol after which they were resuspended in sorbitol (1x of the original volume) and fixed in 50% ethanol (final concentration). The cells were washed and resuspended in 1M sorbitol. DAPI was then added to the solution to a final concentration of 1µg/ml and incubated for 5-10 min at room temperature.

Cells were imaged using a Nikon TE-2000-E inverted microscope with a Apo 40x/NA 0.95 objective and Cascade II 512 camera (Photometrics), Lambda LS Xenon light source and Lambda 10-3 filter wheel control (Sutter Instrument) with a motorized stage (Prior Scientific). All imaging and parameters were set via the Nikon NIS Elements Imaging Software. Images were captured at 1 frame per second through a 89000ET filter set (Chroma Technology) with channels “DIC L”, “FITC” (Ex 490/20, Em 525/36). GFP fluorescence images were collected with an exposure time of 1s. DAPI images for the humanized septin yeast strains were obtained using a Zeiss LSM 710 confocal

microscope with a Plan-Apochromat 63x/1.4 oil-immersion objective with standard DAPI (Ex 350/50, Em 455/50) wavelength setting and operated using Zeiss ZEN Microscope software.

Cell size measurements were performed with a minimum of 10 fields of view and 2000 cells per strain. Image analysis and quantification was performed using FIJI/ImageJ (Schindelin *et al.* 2012, 2015). For quantifying cell size, edge detection scripts were written as Python scripts and FIJI macros.

3.3.6. Mating assays

Since the genes assayed via CRISPR were tested in BY4741 (genotype MATa *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*), the humanized strains were mated to a BY4742 derivative (genotype MATα *his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 mkt1*-D30G *rme1*-ins308A tao3-E1493Q) so that diploids could be selected on SC-Lys-Met medium. Strains were first patched on a YPD plate. The humanized strain being assayed and its complementary mating strain were streaked perpendicular to each other and were grown overnight. These were then replica plated on SC-LYS-MET media to select for diploids. In cases of humanized septin mating assays, haploids were mated on YPD as described above and streaked on SC-Lys-Met to select for diploids following which they were incubated for 3 days before imaging (Fig. 3.6.5B, S7B). For sporulation, diploids were grown overnight on GNA rich pre-sporulation medium. Individual colonies from this plate were inoculated into a liquid sporulation medium containing 0.1% potassium acetate, 0.005% Zinc acetate, and were incubated at vigorous shaking at 25˚C for 3-5 days. Spores were dissected on a YPD plate using a similar protocol, as described previously. Dissected

spores were replica plated on SC-Lys and SC-Met media to select haploid spores. This assay enabled subsequent mating of humanized haploids to each other and also generating homozygous diploids for the human genes *via* CRISPR-Cas9. All heterozygous diploids were scored for spore viability and 2:2 segregation of LYS, MET, and MAT loci.

3.3.7. Gene tree construction

Maximum likelihood trees for each cytoskeletal family were constructed with the Gamma LG protein model with the rapid hill-climbing algorithm bootstrapping 1000 replicates on human and yeast protein sequences curated from the Uniprot database. All trees were computed using Geneious v10.2.6's RAxML's plugin (Stamatakis 2014) (v8.2.11).

3.3.8. Human cytoskeleton RNA expression analysis

RNA expression profiles across human tissue types were obtained from the Human Protein Atlas (Uhlén *et al.* 2015). We extracted the human cytoskeletal gene set from the available expression data normalized via FPKM. The normalized data was then processed in Rstudio and plotted with ggplot2 using the ggridges package.

3.3.9. SCMD-SGD database searching and evaluations

Genetic interaction data for each assayed yeast cytoskeletal gene was downloaded from the Saccharomyces Genome Database (Cherry *et al.* 1998, 2012) (SGD) to curate all the interaction partners for the assayed yeast cytoskeletal gene set. Imaging data and morphology parameter files were downloaded from the Saccharomyces Morphological Database (Saito *et al.* 2004) (SCMD) website. As relevant parameters to the phenotypes

observed in the case of actin and septin interactors, we specifically examined the long (C103) and short axis (C104) lengths of mother cells budded (SCMD cell types B and C) and unbudded cells (SCMD cell type A). C103/C104 ratios were calculated and ranked in ascending order. We reasoned that deletion strains resulting in small cells with round phenotypes (similar to our humanized actin yeast strains) would have ratios close to 1. In contrast, abnormally shaped cells (similar to our humanized septin yeast strains) would have larger (>1) C103/C104 ratios. We ranked the deletion collection phenotypes by the long to short axis ratios of the budded cells as a measure of circularity. Significance was determined by the hypergeometric test.

3.4. RESULTS

3.4.1. Human cytoskeletal genes can functionally replace their corresponding yeast orthologs

Of the 324 cytoskeleton genes in yeast, 101 are essential for growth in standard laboratory conditions and possess identifiable human orthologs as determined by EggNOG (Huerta-Cepas *et al.* 2016) and InParanoid (Sonnhammer and Östlund 2015) (**Fig. 3.6.S1**). We focused on orthologs constituting major structural elements of the eukaryotic cytoskeleton, including actin, myosin, septin, and tubulin gene families identifying 106 testable human-yeast ortholog pairs. By restricting our tests to elements of the yeast cytoskeleton essential for cell growth in standard lab conditions, we could initially assay replaceability of human genes in yeast *via* simple growth rescue assays. Since human cytoskeletal gene families have undergone multiple duplication events, we systematically swapped each human gene within a family in place of its corresponding

yeast ortholog to assay if present day human genes could still complement the lethal loss of their yeast ortholog(s).

Complementation assays in yeast were carried out in 3 different ways (**Fig. 3.6.2**), wherein orthologous yeast genes could be (i) genetically segregated away after sporulation of a heterozygous diploid deletion strain (Wach *et al.* 1994; Winzeler *et al.* 1999; Giaever *et al.* 2002) (**Fig. 3.6.S2A**) (ii) inactivated *via* a temperature-sensitive allele(Li *et al.* 2011; Kofoed *et al.* 2015) (**Fig. 3.6.S3A**) or (iii) endogenously replaced by homologous recombination repair with the human ortholog(s) following CRISPR-Cas9 cleavage (Kachroo *et al.* 2017; Akhmetov *et al.* 2018a) (**Fig. 3.6.S4A**).

Each assayed human gene was sequence verified and either sub-cloned into a single-copy (CEN) yeast expression vector transcriptionally controlled by a constitutive GPD promoter [in assays (i) and (ii)] or PCR amplified with flanking homology to the yeast locus of interest [in assay (iii)]. In cases involving heterozygous diploid sporulation assays, we leveraged the heterozygous diploid deletion collection (Giaever *et al.* 2002), wherein one copy of the yeast ortholog being queried has been replaced with a KanMX resistance module thus enabling selection of haploids with either the null (conferring G418 antibiotic resistance) or the wild-type allele (susceptible to G418) post sporulation (**Fig. 3.6.S2B, C**). In cases of temperature-sensitive (ts) haploid strains, a single gene of interest is modified to carry mutations inactivating its corresponding gene product at restrictive temperatures (37°C) but permitting growth at lower temperatures (25°C), thus allowing for temperature-dependent growth rescue complementation assays (**Fig. 3.6.S3B, C**). Finally, in cases of endogenously chromosomal replacement, human genes

were assayed by natively substituting their yeast ORF mediated *via* CRISPR-Cas9 (Akhmetov *et al.* 2018a) (**Fig. 3.6.S4B, C**).

By taking advantage of full-length human cDNA clones from the human ORFeome(Rual *et al.* 2005; Lamesch *et al.* 2007; Temple *et al.* 2009) and existing yeast strains with null (or conditionally null) alleles for the relevant orthologs (Winzeler *et al.* 1999; Giaever *et al.* 2002; Li *et al.* 2011; Kofoed *et al.* 2015), we could successfully assayed 89 of 109 human-yeast ortholog pairs, accounting for 50 of 62 (~81%) human genes across the 7 major cytoskeletal gene families. We observed that 13 (~26%) of the 50 tested human genes could functionally replace their yeast orthologs in at least one of the three assay types, while 37 (~72%) could not (**Fig. 3.6.2**). In particular, we found that 2 of 7 actin, 4 of 14 myosin (heavy chain), 4 of 13 septin, 2 of 9 β- and 1 of 2 ɣ-tubulin human genes complemented lethal growth defects caused by loss of their corresponding yeast orthologs, whereas genes from the human light chain myosin and α-tubulin families did not (**Fig. 3.6.3**). Since the septin family had expansions in both human and yeast lineages, we systematically assayed all human septins against 4 yeast deletion backgrounds (*CDC3*, *CDC10*, *CDC11*, and *CDC12*), observing that complementing human septins functionally replaced only *CDC10*. Thus, within 5 of these 7 essential cytoskeletal gene families, at least one extant human gene could substitute for its yeast ortholog, indicating that the yeast and human genes both still executed the essential roles of their shared opisthokont ancestor.

To assay robustness of complementation and better measure the extent to which human orthologs complemented in standard laboratory conditions, we performed quantitative growth assays on the humanized strains. We measured growth at 30°C

(where possible, temperature-sensitive strains were measured at 37°C) (**Fig. 3.6.S5**). Humanized strains assayed *via* plasmid-based complementation of the heterozygous diploid sporulation assay were assayed both in selective (SC-Ura) and rich medium (YPD). These strains while retaining the human gene encoding plasmid showed no major growth defects in YPD. However, they exhibited significantly slower growth rates in SC-Ura medium compared to their wild-type yeast counterparts (**Fig. 3.6.S5A**). Besides Hs-*ACTBL2* and Hs-*SEPT3*, all other humanized strains showed drastically reduced growth rates in SC-Ura selective medium (**Fig. 3.6.S5A (i)**) compared to wild-type yeast. In particular, we found that strains with humanized heavy myosins and septins had mean doubling times over twice as long as wild-type (**Fig. 3.6.S5A (i), Table 3.7.1**). However, in rich medium (YPD) these fitness defects were substantially rescued (**Fig. 3.6.S5A (ii)**). We found that replacing Hs-*MYH9* particularly affected growth to saturation phase in YPD with doubling time in these strains being twice the wild-type rate (**Fig. 3.6.S5A (ii), Table 3.7.1**). With respect to temperature sensitive growth rescue assays we found that humanized strains assayed in this manner generally exhibited slower growth rates with humanized ɣ-tubulin (Hs-*TUBG2*) strains in particular showing drastically slow growth rates, doubling at one-quarter the rate of wild-type lab strains and failing to reach comparable biomass at saturation (**Fig. 3.6.S5B, Table 3.7.1**). In contrast, we genomically replaced β-tubulins Hs-*TUBB4* and *Hs-TUBB8* both shared similar growth profiles to wild-type yeast strains (**Fig. 3.6.S5C, Table 3.7.1**).

To probe growth effects of humanization in more detail, we subjected the strains (except those assays performed in a temperature-sensitive strain background) to temperature stress, repeating the growth assays at low and high temperatures. While most

humanized strains grew at wild-type rates at 25°C (**Fig. 3.6.S11 bottom**), we observed differential effects at 37°C (**Fig. 3.6.S11 top**). Specifically, yeast with human septins Hs-*SEPT6*, Hs-*SEPT9*, and Hs-*SEPT10* all reached stationary phase at a lower biomass compared to the wild-type yeast. Although septin genes complemented the lethal growth defect, they performed sub-optimally, especially under stress conditions, suggesting that they may be failing to complement other non-essential roles of the yeast orthologs, meriting a deeper examination of additional gene family specific phenotypes, especially in light of recent observations that *cdc10Δ* phenotype severity may also be modulated by altering plasma membrane properties (Michel *et al.* 2017).

3.4.2. Sporulation and meiotic roles of human β-tubulin are regulated by their native regulation in yeast

In yeast, β-tubulin (*TUB2*) is primarily associated with chromosome segregation during budding (Neff *et al.* 1983; Vogel *et al.* 2001; Bode *et al.* 2003) (asexual reproduction involving mitosis), mating (Bardwell 2005; Molk and Bloom 2006) (sexual reproduction) and sporulation (Neiman 2011) (starvation response involving meiosis). In our complementation assays, we observed that the human β-tubulins Hs-*TUBB4* and Hs-*TUBB8* only complemented the loss of *TUB2* when genomically inserted at the yeast *TUB2* locus but failed to do so in plasmid-based segregation assays in heterozygous diploid knockout strains. Notably, plasmid expression was controlled transcriptionally by a constitutive GPD promoter; both human genes failed to rescue its corresponding heterozygous diploid yeast gene deletion leading to sporulation defects producing either mis-segregated (**Fig. 3.6.4A - left panel**) or inviable spores (**Fig. 3.6.4B - left panel**). Therefore, replacing endogenous *TUB2* with human β-tubulins Hs-*TUBB4* and Hs-

TUBB8 at a minimum supported mitosis and asexual cell division *via* budding, but it was unclear if they supported meosis and sporulation.

In order to determine whether replaceable human β-tubulins also rescued meiotic and sporulation specific roles of their yeast ortholog, we mated haploid humanized strains (harboring human genes at the corresponding yeast genes' native genomic loci) with wild-type yeast. These humanized strains successfully mated with wild-type yeast strains to produce viable diploids. We assayed if these heterozygously humanized strains could successfully perform meiosis by sporulating the diploids generated from our mating assays. We observed high spore viability for both Hs-*TUBB4* (~96.9%) and Hs-*TUBB8* (~84.4%) (**Fig. 3.6.4A and 4B - right panel**). Having shown that both replaceable β-tubulins can mate with the wild-type strains (yielding high sporulation efficiency similar to the wild-type diploids) (**Fig. 3.6.S6A - right panel)**, we next asked whether strains homozygous for human β-tubulins could behave similarly. We sporulated the humanized homozygous β-tubulin diploid strains assaying for proper progression through meiosis *via* sporulation and found that these diploid humanized yeast strains also sporulated in both Hs-*TUBB4* (93.8%) and Hs-*TUBB8* (84.4%) cases. **(Fig. 3.6.4A and 4B - middle panel)** similar to their heterozygous (**Fig. 3.6.4A and 4B - right panels**) and wild-type diploids (~87.5%) (**Fig. 3.6.S6A - right panel** including 2:2 segregation of the MAT, LYS and MET loci after dissecting tetrads, see **Fig. 3.6.S6**). These results reveal that native regulation of β-tubulin is more critical to yeast than the species origin of the gene. The functional replacement of *TUB2* by two human orthologs, Hs-*TUBB4* and Hs-*TUBB8*, requires insertion into the native yeast locus, but nonetheless enables these genes

to successfully perform their roles in meiosis and sporulation successfully, similar to the yeast β-tubulin *TUB2* (**Fig. 3.6.S6)**

3.4.3. Human septin orthologs and their isoforms can carry out *CDC10*'s meiotic and mating role

Next, we considered if human orthologs of yeast septins could also perform conditionally essential cytoskeletal roles in yeast. The yeast septin family is a particularly interesting case of gene family expansion as there have been duplications in both the human and yeast lineages, suggesting ancestral functions might have been distributed across paralogs in both species. Yeast possess 7 septin genes (*CDC3*, *CDC10*, *CDC11*, *CDC12*, *SPR3*, *SPR28*, and *SHS1*) while humans have 13 septin genes(Pan *et al.* 2007) (**Fig. 3.6.1, 3**). 3 of the 7 yeast septins (*CDC3*, *CDC11*, and *CDC12*) are essential for vegetative growth (budding) in standard lab conditions. We performed 39 of the 91 total possible septin complementation assays in each of the 3 septin null backgrounds, we found that none of the human orthologs assayed could complement *CDC3*, *CDC11*, and *CDC12*.

While the remaining 4 septins (*CDC10*, *SPR3*, *SPR28*, and *SHS1*) are not essential for vegetative growth, *CDC10* however, plays essential roles in meiotic cell division, mating, and spore health, with *cdc10Δ* strains exhibiting severe mating and sporulation defects (Douglas *et al.* 2005; McMurray *et al.* 2011; Neiman 2011; Kim and Rose 2015). Indeed, we also observed that the diploid strain heterozygously null for *CDC10* sporulated poorly, almost always producing only 2 viable spores (**Fig. 3.6.S2B(ii)**).

To identify human orthologs capable of rescuing sporulation defects caused by the loss of *CDC10,* we systematically assayed each of the 13 human septin orthologs. We observed that 4 of 13 septins (Hs-*SEPT3*, Hs-*SEPT6*, Hs-*SEPT9*, Hs-*SEPT10*) were capable of functionally complementing *CDC10*'s meiotic and sporulation roles (**Fig. 3.6.3**). Since *cdc10Δ* strains additionally show severe mating defects(McMurray and Thorner 2008; McMurray *et al.* 2011) (**Fig. 3.6.S7A**), we next asked if the expression of individual human septin orthologs in a *cdc10Δ* strain could facilitate mating. We assayed this by cloning human septin genes into yeast expression vectors (**Materials and Methods**) and individually transformed *cdc10Δ* strains with each human septin clone. Subsequently, we mated each *MATa* transformant to a complimentary *MATα* strain (**Fig. 3.6.5A**). While only 4 of 13 tested human septins, Hs-*SEPT3*, Hs-*SEPT6*, Hs-*SEPT9*, and Hs-*SEPT10*, rescued *CDC10*'s essential meiotic role, we found that all assayed septins were capable of performing *CDC10's* role in mating (**Fig. 3.6.5B**).

Unlike in yeast, human septins are known to exhibit a variety of different splice forms (Weems and McMurray 2017; Hall and Russell). Based on the availability of verified splice form (as per the Uniprot database) in the human ORFeome, we tested different isoforms of Hs-*SEPT6* and Hs-*SEPT10* observing contrasting results with both sets of isoforms. While the canonical isoform of Hs-*SEPT6* functionally complemented its yeast ortholog *CDC10*, its testes isoform with an extended C-terminus failed to do so (**Fig. 3.6.5C**). However, when testing the uterine and canonical isoforms of Hs-*SEPT10*, the opposite was true, with only the truncated uterine form functionally complementing *CDC10* (**Fig. 3.6.5D**). Thus for both, the presence of an extended C-terminus negatively affected the ability of the tested human septin isoforms to replace. Taken together, these

results suggest that human septin orthologs can effectively rescue essential *CDC10* phenotypes in mating and meiosis.

3.4.4. Humanized yeast strains differ in cell morphology

Eukaryotic cytoskeletal proteins dynamically control cell shape and morphology. We next asked whether humanizing elements of the yeast cytoskeleton would result in any visually obvious phenotypic changes to cell morphology. Using light and fluorescence microscopy, we imaged the 13 humanized strains generated and quantified cell shape and size across >22,000 individual cells (**Materials and Methods**). In the case of humanized heavy myosins, Hs-*MYH7B* and Hs-*MYH11*, we did not observe any obvious defects in the cell morphologies, but introducing Hs-*MYH1* and Hs-*MYH9* induced a slight size increase leading to more spherical cells (**Fig. 3.6.6A, S7**). However, the effect, while significant, was small, and their median sizes were within 1% of wild-type (**Fig. 3.6.6A, S8**). In contrast, strains with humanized actins, tubulins, and septins had visibly different cell morphologies (**Fig. 3.6.6A, S8**). Humanizing actin visibly reduced cell size, resulting in round/spherical cells with small buds (**Fig. 3.6.6A, S8**). Complementing human ɣ-tubulin Hs-*TUBG2* showed the opposite phenotype, resulting in enlarged and ovoid cells (**Fig. 3.6.6A, S8**). The endogenously replaced human β-tubulin, Hs-*TUBB4*, also showed reduced cell size (**Fig. 3.6.6A, S8**). Since α- and β-tubulins physically associate to form heterodimers we introduced human β-tubulins Hs-*TUBB4* and Hs-*TUBB8* into a strain expressing GFP-tagged α-tubulin (Huh *et al.* 2003) (*TUB1-GFP*), allowing us to visualize microtubules and observe if chimeric yeast-human αβ-tubulin heterodimers were being assembled (**Fig. 3.6.S10A**). While we observed microtubules in both strains, we noticed different cell morphologies when humanizing

Hs-*TUBB4* (**Fig. 3.6.S10A**), indicating a synthetic genetic interaction between the GFP tagged form of *TUB1* and Hs-*TUBB4*, absent from the interaction of Hs-*TUBB4* and native *TUB1*. Examining the distributions further, we observed a bimodal distribution of cell sizes (**Fig. 3.6.S10B**) for Hs-*TUBB4* suggestive of cell cycle defects (as for *TUB2* Ser172 mutations (Caudron *et al.* 2010)). In contrast, the cell morphology and microtubule assembly of the Hs-*TUBB8 TUB1-GFP* strain appeared similar to wild-type.

All replaceable septin strains (Hs-*SEPT3*, Hs-*SEPT6*, Hs-*SEPT9*, and Hs-*SEPT10*) showed abnormal cell morphologies (**Fig. 3.6.6A, B, S8**), notably reminiscent of elongated pseudohyphal forms often observed in septin mutant strains as a function of improper activation of the morphogenesis checkpoint in the budding yeast cell cycle (Keaton and Lew 2006; McMurray *et al.* 2011). In order to verify if the replaceable human septins localized to the bud neck, we expressed C-terminal EGFP-tagged Hs-*SEPT4* (non-replaceable) and Hs-*SEPT9* (replaceable) in heterozygous diploid deletion strains. We observed that when sporulated, the *eGFP*-tagged, Hs-*SEPT9,* can also rescue the sporulation defect induced by deletion of *CDC10* (**Fig. 3.6.S9A**), whereas the mating defect in *cdc10Δ* haploids is rescued by both Hs-*SEPT9* and Hs-*SEPT4 eGFP*-tagged genes (**Fig. 3.6.S9B**). We then tested the localization of both these human septins in haploid *cdc10Δ* strains. We did not see the localization at the bud neck (**Fig. 3.6.S9C**); rather, we observed the GFP fluorescence diffused throughout the cell. Therefore, we conclude that the likely mode of the rescue of yeast septin function is partial or indirect in nature. The assay also reveals that the strains do not have any major spore formation defect as we see several colonies growing on SC-Arg-His-Leu+Can-Ura-G418 medium similar to the empty vector control transformed strain (**Fig. 3.6.S9A**). While it has been

reported that *cdc10*Δ null mutants do have aneuploid genomes (McMurray *et al.* 2011), in the conditions that we tested (heterozygous diploid deletion sporulation assay in SC-Arg-His-Leu+Can-Ura+/-G418 with growth for 3 days), we rule out aneuploidy as a consequence of haploinsufficiency. However, we also observed that many *cdc10Δ* haploid colonies that grow in the presence of human gene (Hs-*SEPT9*) were able to also grow independent of the plasmid suggesting that the replaceability is not optimal and the expression of Hs-*SEPT9* induces segregation defects in these cells that may likely induce aneuploidy (**Figure 3.S9A**).

Taken together, these data suggest that human septins fail to fully complement the role of *CDC10*. The fraction of elongated cells differed across human septins, with Hs-*SEPT3* producing lower proportions of elongated cells (**Fig. 3.6.6A, S8**). To assay whether this effect arose from defective cytokinesis, we quantified the nuclei per cell by DAPI staining. We indeed observed that the cells were multinucleated (**Fig. 3.6.6B**). Despite this severe cell morphology defect, all replaceable human septins still enabled the strains to maintain growth rates comparable to wild-type strains (**Fig. 3.6.3D**). Taken together, we found that multiple human septins can rescue the essential meiotic and segregation roles of *CDC10*, but result in abnormal cell morphologies with delayed and/or defective cytokinesis.

3.4.5. Humanized cytoskeletal orthologs phenocopy cell morphology defects observed by deleting interaction partners

In spite of rescuing lethal growth defects, the humanized strains showed visible cell morphology defects consistent with the known roles of the humanized genes. These defects suggested that the complementing human genes might be failing to perform some

of the non-essential roles of the yeast cytoskeletal orthologs regulating cell shape and morphology. While the replaceable human cytoskeletal genes functionally complement the lethal growth defect caused by deletion of their yeast orthologs, we hypothesized that they failed to fully interface with their constituent interaction networks, thereby breaking key non-essential interactions regulating cell morphology. If true, this would imply that the humanization of a particular cytoskeletal yeast gene might phenocopy the deletion of its corresponding non-essential yeast interaction partner(s) (**Fig. 3.6.7A**).

To test this hypothesis, we curated both physical and genetic interaction partners of all humanizable yeast orthologs from the Saccharomyces Genome Database (SGD) (Cherry *et al.* 1998, 2012) and mined the Saccharomyces Cerevisiae Morphological Database (SCMD) (Saito *et al.* 2004), a database cataloging the morphologies of ~1.9 million cells across 4,718 haploid non-essential gene deletion backgrounds in yeast and measuring ~ 501 cell shape parameters. Specifically, we considered the humanized strains showing drastic morphology changes and computed the ratio of long to short cell axes as a measure of cell roundness or elongation. Deletion strains with elongated phenotypes (*i.e.*, similar to humanized septin strains) would have higher axes ratios, and strains with round/spherical cells (*i.e.*, similar to humanized actin strains) would have axes ratios centered around 1. To ensure stringency in curating interactors of both *ACT1* and *CDC10*, we restricted the interactions *a priori* to those reported at least twice in the literature. Additionally, we further restricted the interacting partners list to only those that contained SCMD cell morphology measurements for the corresponding deletion strain. To test our hypothesis, we next plotted the z-scores of the circularity index and asked whether there was a significant enrichment of *ACT1* interactors in deletion strains with

circular morphologies (<2 standard deviations) and contrastingly an enrichment of *CDC10* interactors in deletion strains with elongated morphologies (>2 standard deviations) (**Fig. 3.6.7B**). Indeed, we found a significant overlap, albeit not a very large one, between both the physical and genetic interactors of humanizable cytoskeletal genes and their observed phenotypes when deleted. In particular we found 3 physical ($p \leq 8.9 \times 10^{-4}$) and 11 genetic interactors ($p \leq 2.9 \times 10^{-6}$) of *ACT1* enriched for drastic cellular morphologies. In the case of yeast septins, we found 3 physical ($p \leq 4.3 \times 10^{-4}$) and 10 genetic ($p \leq 5.9 \times 10^{-6}$) interactors of *CDC10* enriched for elongated morphologies (**Fig. 3.6.7B & C**). Taken together, these data suggest that the replaceable human orthologs may indeed fail to maintain non-essential physical interactions controlling cell shape and morphology and genetically perturb relevant morphology genes. It is possible that the amplification of cytoskeletal elements in the human lineage may have led to distribution of specific sub-functions and/or loss of interactions across duplicated human genes compared to the opisthokont ancestor. Our observations suggest that replaceable human cytoskeletal genes functionally complement the essential roles of their yeast ortholog, but simultaneously break key non-essential cytoskeletal associations, thereby phenocopying the deletion of their interaction partners.

3.5. DISCUSSION

Identifying how genes retain and/or distribute roles across their families is key to understanding the diversification of conserved genes across organisms. Cross-species gene swaps provide an opportunity to directly test the functional divergence of

orthologous genes even over large evolutionary timescales. While *Saccharomyces cerevisiae* (Baker's yeast) and *Homo sapiens* (humans) diverged from their opisthokont ancestor nearly a billion years ago, both species still share thousands of orthologous genes and high-throughput humanization assays in yeast have found that many human genes are capable of substituting for their yeast orthologs with rates up to 47%, depending on strains and assays (Kachroo *et al.* 2015; Hamza *et al.* 2015; Laurent *et al.* 2016, 2019; Sun *et al.* 2016; Yang *et al.* 2017). Most of these complementation tests until recently were performed in the absence of gene family expansions and revealed many humanizable systems, including the proteasome, and sterol and heme biosynthesis pathways (Kachroo *et al.* 2015, 2017; Hamza *et al.* 2015; Sun *et al.* 2016; Yang *et al.* 2017; Laurent *et al.* 2019). In this study, we sought to better understand the functional equivalence of human and yeast orthologs that play key structural roles in the eukaryotic cytoskeleton, particularly focusing on how gene family expansions in cytoskeletal lineages might have diversified in function across their respective gene families.

By systematically humanizing key structural components central to the yeast cytoskeleton, we determined that 5 of the 7 (~71%) assayed eukaryotic cytoskeletal gene families could be successfully humanized by at least one human ortholog within a family. In all, 13 (26%) out of 50 tested human cytoskeletal proteins could at least partially substitute for the corresponding yeast gene and complement a lethal growth defect caused upon loss of the yeast ortholog (**Fig. 3.6.3A**). Interestingly, these results are broadly consistent with our previous study, where 17 (~28%) of 60 tested regulatory genes of the human cytoskeleton replaced their yeast orthologs (Laurent *et al.* 2019) (**Fig. 3.6.S1**).

It is noteworthy that we were unable to successfully humanize yeast with any member of the myosin light chain or α- tubulin gene families, suggesting extensive functional divergence in yeast-human ortholog pairs resulting in failure to substitute for each other across species. This variation in replaceability was not merely explainable by obvious patterns across human genes (at least as captured by the variation in their expression patterns across tissues; **Fig. 3.6.S12A**), nor was it explained by their degree of sequence similarity. As with previous yeast humanization studies (Kachroo *et al.* 2015), sequence conservation among replaceable and non-replaceable human cytoskeletal genes did not significantly predict replaceability (**Fig. 3.6.S12B**). While uncovering additional new properties predicting replaceability is beyond the scope of this study, future efforts at systematically constructing chimeric human/yeast genes have the potential to reveal which regions of the human/yeast orthologs are critical to maintain functional compatibility, perhaps enabling targeted humanization of specific domains and regions of genes (Barger *et al.* 2019).

With only slight differences in mitotic growth between wild-type and most humanized strains observed under standard laboratory growth conditions, our results suggest robust complementation of essential cellular roles in most cases. However, we subsequently found incomplete complementation of multiple non-essential cellular roles. We found that human β-tubulins Hs-*TUBB4* and Hs-*TUBB8* remarkably complemented *TUB2*'s roles even in sexual reproduction, including sporulation and mating. However, this was contingent on genomic integration under the native yeast regulation, consistent with previous studies showing that *TUB2* expression levels are tightly regulated (Beilharz

et al. 2017) with overexpression of β-tubulin leading to toxicity, chromosome loss and cell cycle arrest (Burke *et al.* 1989; Weinstein and Solomon 1990; Abruzzi *et al.* 2002).

For the septin gene family, which has expanded to consist of 7 yeast genes to 13 in humans, we found evidence for functional divergence across family members in both human and yeast lineages, with human genes tending to only fulfill a subset of specific roles of their yeast ortholog(s). While none of the human septin orthologs individually rescued essential roles of the yeast *CDC3*, *CDC11*, and *CDC12* genes, we found that 4 human septin orthologs (Hs-*SEPT3*, Hs-*SEPT6*, Hs-*SEPT9*, Hs-*SEPT10*) complemented *CDC10*'s essential roles in meiosis and sporulation. However, severe mating defects caused by deleting *CDC10* in a haploid strain background were completely rescued by every assayed member of the human septin family (**Fig. 3.6.5A, B**), indicating that all human septin orthologs executed the roles of their yeast counterparts to at least some extent. Extending this theme further, we saw that replaceability of Hs-*SEPT6* and Hs-*SEPT10* differed based on their protein isoforms, demonstrating functional divergence not just across human septin paralogs but also within splice forms of a gene (**Fig. 3.6.5C, D**). Taken together, such specific complementation patterns suggest a complex evolutionary trajectory and delegation of function across the septin gene family in eukaryotes. It remains to be seen if human septin orthologs can individually complement their non-essential yeast counterparts *SHS1*, *SPR3*, and *SPR28*. While septin gene family expansions have been understood to predominantly bring functional redundancy and robustness within its interactome, recent studies have identified tissue specific roles within human septin orthologs (Fujishima *et al.* 2007; Tsang *et al.* 2008; Kaplan *et al.* 2017). The mechanical roles facilitated by human and yeast septin gene families appear

to be remarkably conserved albeit their involvement in seemingly unrelated processes across species (Falk *et al.* 2019).

Over the course of these complementation assays, we observed diverse cell morphologies among humanized strains that were broadly gene family specific, with complementing myosins and β-tubulins largely remaining unchanged, but actin, ɣ-tubulin, septin families showing characteristic morphological differences as drastic as defective cytokinesis. Owing to the crucial role of the cytoskeleton in maintaining cell shape, we found that replaceable human cytoskeletal genes tend to perform its yeast ortholog's essential roles equivalently while simultaneously breaking key non-essential cytoskeletal genetic interactions regulating cell morphology. In agreement with our findings, a recent study systematically determining cell size regulators in yeast found that 145 genes of ~400 deletion strains were genetic interactors of actin (Soifer and Barkai 2014). Replaceable human septin orthologs, in particular, represent interesting and rather extreme cases of abnormal cell morphology. Previously, a study showed that introducing the *A. nidulans* septin *AspC* in *S. cerevisiae* induces a similar pseudohyphal morphology observed when substituting for *CDC12* in the septin ring (Lindsey *et al.* 2010). More recently, a report (Kim and Rose 2015) demonstrated that doubly deleting *ELM1* and *FUS3* (both genetic interactors of *CDC10*) in yeast produces the same filamentous elongated morphology with similar cytokinesis defects observed when humanizing *CDC10*. In agreement with these studies, our results suggest that human septins sufficiently complement *CDC10'*s roles by surpassing cellular thresholds for growth but not morphology. While in cases of both yeast actin and septin, we indeed find significant enrichment of their corresponding genetic interactors leading to abnormal morphologies,

it is difficult to mechanistically interpret the consequences of genetic interactions or lack thereof. However, with significant enrichment in the physical interactors of *ACT1* and *CDC10* associated with the observed cell morphologies, our analyses suggest that human cytoskeletal orthologs in yeast may break at least some non-essential protein-protein interactions underlying normal cell morphology. It remains to be seen if humanization of their corresponding interaction partners can revert yeast cells to normal wild-type like morphologies. The genetic and biochemical mechanisms by which human orthologs regulate cytoskeletal interactions in yeast are yet to be explored.

Systematic swaps of humanized cytoskeletal elements in yeast now provide a direct view of how compatible human orthologs likely are within their corresponding yeast interaction network(s), pointing to conserved and divergent interactions among eukaryotes. While our complementation assays tested the ability of human cytoskeletal orthologs to singly complement their yeast equivalents, combinatorial multi-gene swaps might enable humanization of entire yeast systems to study modularity and paralog-level cross-talk between different human cytoskeletal families in a genetically tractable eukaryote. Widening the scope of cytoskeletal humanization efforts to include accessory motors and chaperones, including kinesins and dyneins, could help advance our understanding of eukaryotic cytoskeletal evolution. It remains to be seen if constructing a fully human cytoskeleton in yeast would be feasible.

These humanized strains can now serve as cellular reagents to study complex human cytoskeletal processes in a simplified eukaryotic context, allowing functional roles of distinct family members to be assayed individually. Screening allelic variants and mutational libraries using these strains might enable the rapid identification of disease

variants in a high-throughput manner (Kachroo *et al.* 2015; Hamza *et al.* 2015; Sun *et al.* 2016, 2018; Yang *et al.* 2017; Marzo *et al.* 2019), paving the way for a better understanding of the genetic and molecular basis of cytoskeletal disorders.

3.6. FIGURES

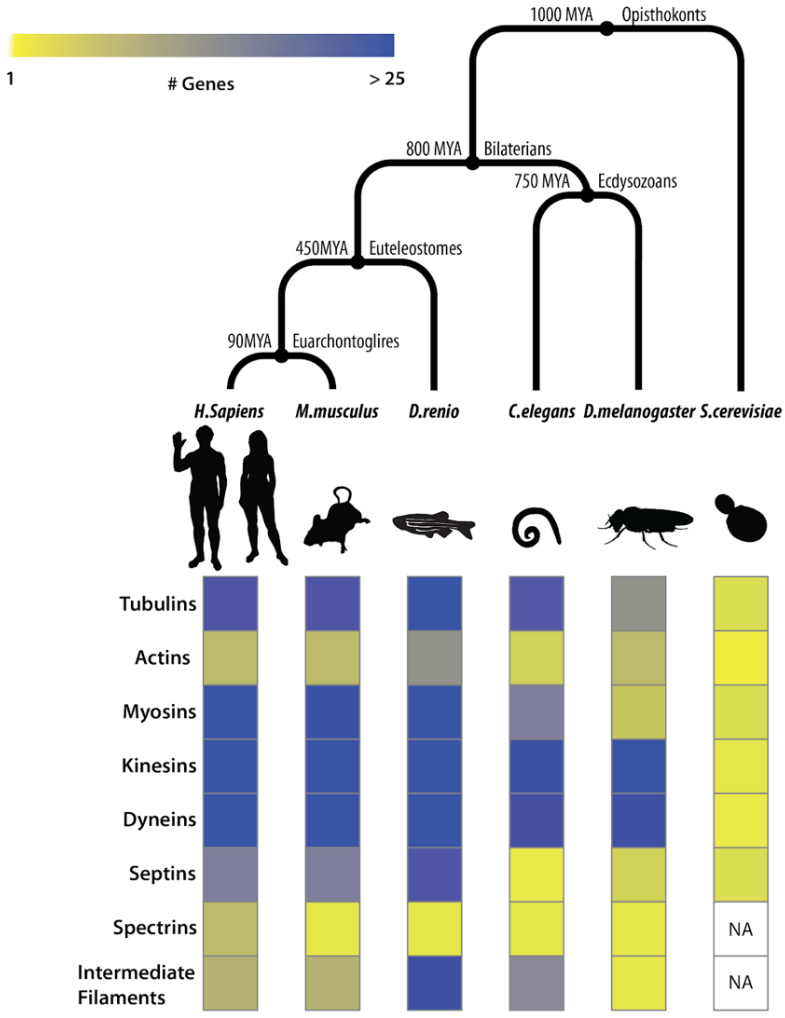


Figure 3.6.1. Orthologs in cytoskeletal gene families have undergone extensive duplications in Bilaterians.

(A) (Top) Species divergence across opisthokonts. (Bottom) Heatmap depicting the number of orthologs in eukaryotic cytoskeleton gene families (rows) across species (columns). Cytoskeletal ortholog counts for model organisms curated from The Alliance of Genome Resources database (Agapite *et al.*).

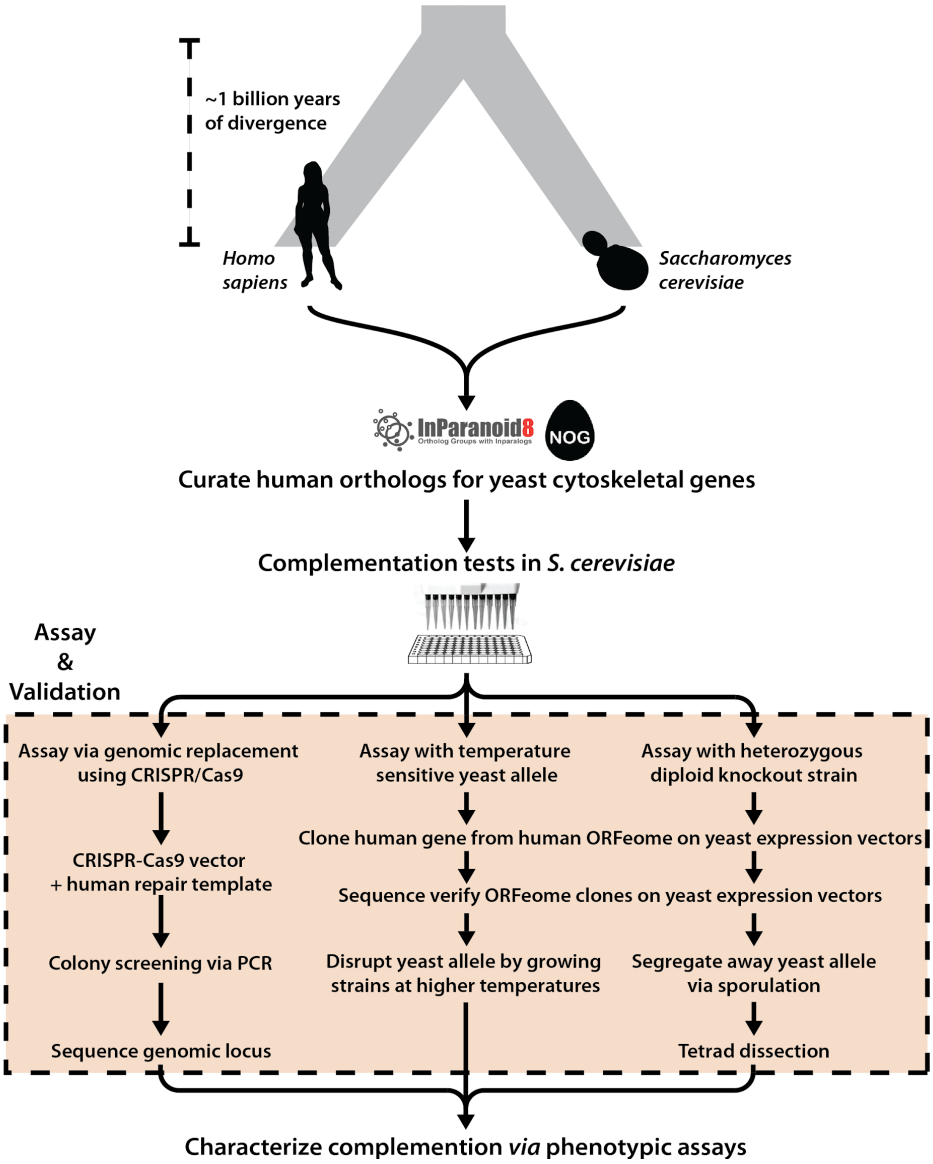


Figure 3.6.2. Overview of humanization assays.

For each human-yeast ortholog pair (curated from inParanoid and EggNOG), complementation assays in *Saccharomyces cerevisiae* were performed using 3 strategies: (i) genomic replacement at the native yeast loci via CRISPR-Cas9, (ii) temperature-sensitive inactivation of the yeast allele, and (iii) sporulation of a heterozygous diploid

deletion strain followed by tetrad dissection. Complementing human orthologs were further characterized using various phenotypic assays, including quantitative growth measurements, environmental stress tests, mating, and segregation assays.

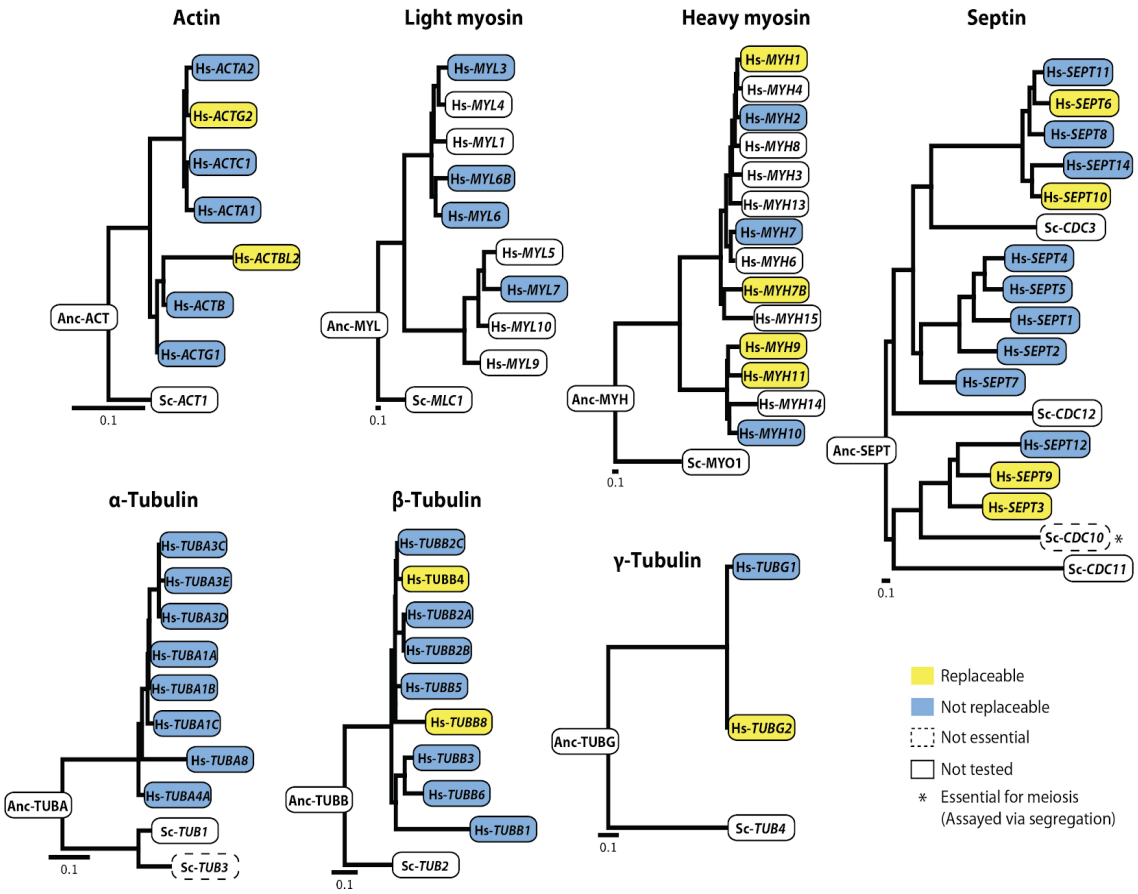


Figure 3.6.3. Human cytoskeletal genes replace their corresponding yeast orthologs.

5 of 7 tested cytoskeletal families possess at least one functionally replaceable human ortholog. Neither of the tested human light myosin or α-tubulin genes could replace their corresponding yeast versions. Each human septin ortholog was tested for replaceability in 4 yeast septin null backgrounds (*CDC3*, *CDC10*, *CDC11*, *CDC12*), however, human septins complemented only *CDC10*, non-essential for mitotic growth but essential in segregation and mating. (Orthology relationships between human genes and yeast septins *SHS1*, *SPR3*, *SPR28* not shown). All phylogenetic gene trees were constructed with

protein sequences of their respective orthologs (see **Materials and Methods**). Scale bars indicate expected substitutions per site.

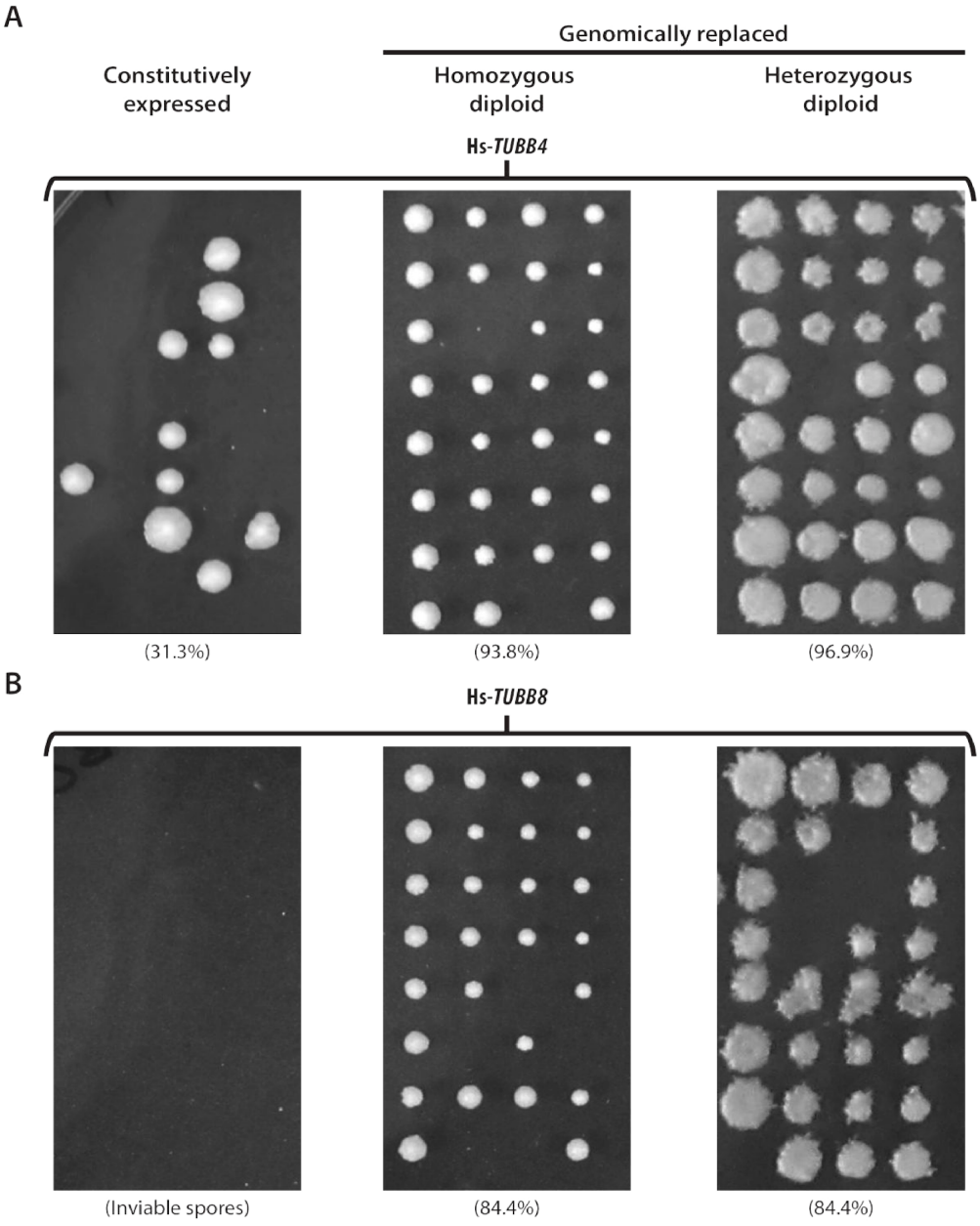


Figure 3.6.4. Replaceability of human β-tubulins Hs-*TUBB4* and Hs-*TUBB8* is determined by their native yeast expression/regulation.

Efficiency of sporulation and segregation of both (A) Hs-*TUBB4* and (B) Hs-*TUBB8* tested in 3 different genetic backgrounds *via* tetrad dissection. Both constitutively

expressed human β-tubulins Hs-*TUBB4* and Hs-*TUBB8* (controlled by a GPD promoter) fail to complement (left panel) whereas when replaced genomically could both grow and mate with wild-type yeast (middle panel). Diploids (both hetero- and homozygous) could also proceed through sporulation and meiosis similar to the wild-type yeast diploid strain (middle and right panel, see **Materials and Methods**). For population level complementation assays of the heterozygous diploid deletion strains see **Fig. 3.6.S6**. Spore viability indicated in brackets.

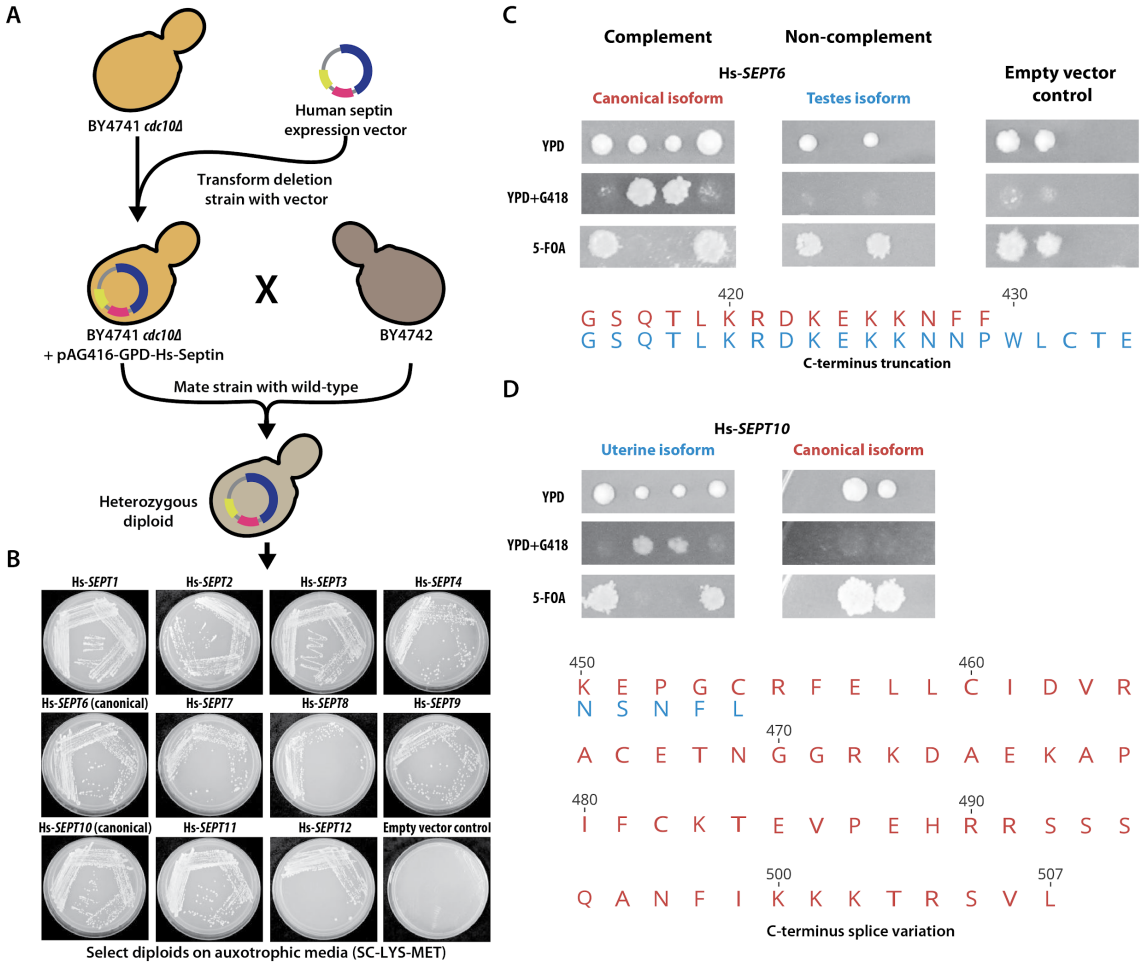


Figure 3.6.5. Human septins differentially rescue meiotic and mating roles of the yeast *CDC10*.

(A) Mating rescue assay for BY4741 *cdc10Δ*. (B) All assayed human septins expressing strains can mate with BY4742, whereas the empty vector containing BY4741 *cdc10Δ* fails to rescue the mating defect caused by deleting *CDC10*. Plates were imaged after 3 days of incubation on SC-Lys-Met (C) and (D) depict differential replaceability of human septin splice forms, Hs-*SEPT6* and Hs-*SEPT10*, respectively. The top panels demonstrate the heterozygous diploid deletion mutant segregation assay for the assayed isoforms

while the bottom panels show the sequence alignment of the variation across the canonical (red) and tissue-specific (blue) isoforms tested for both Hs-*SEPT6* and Hs-*SEPT10*.

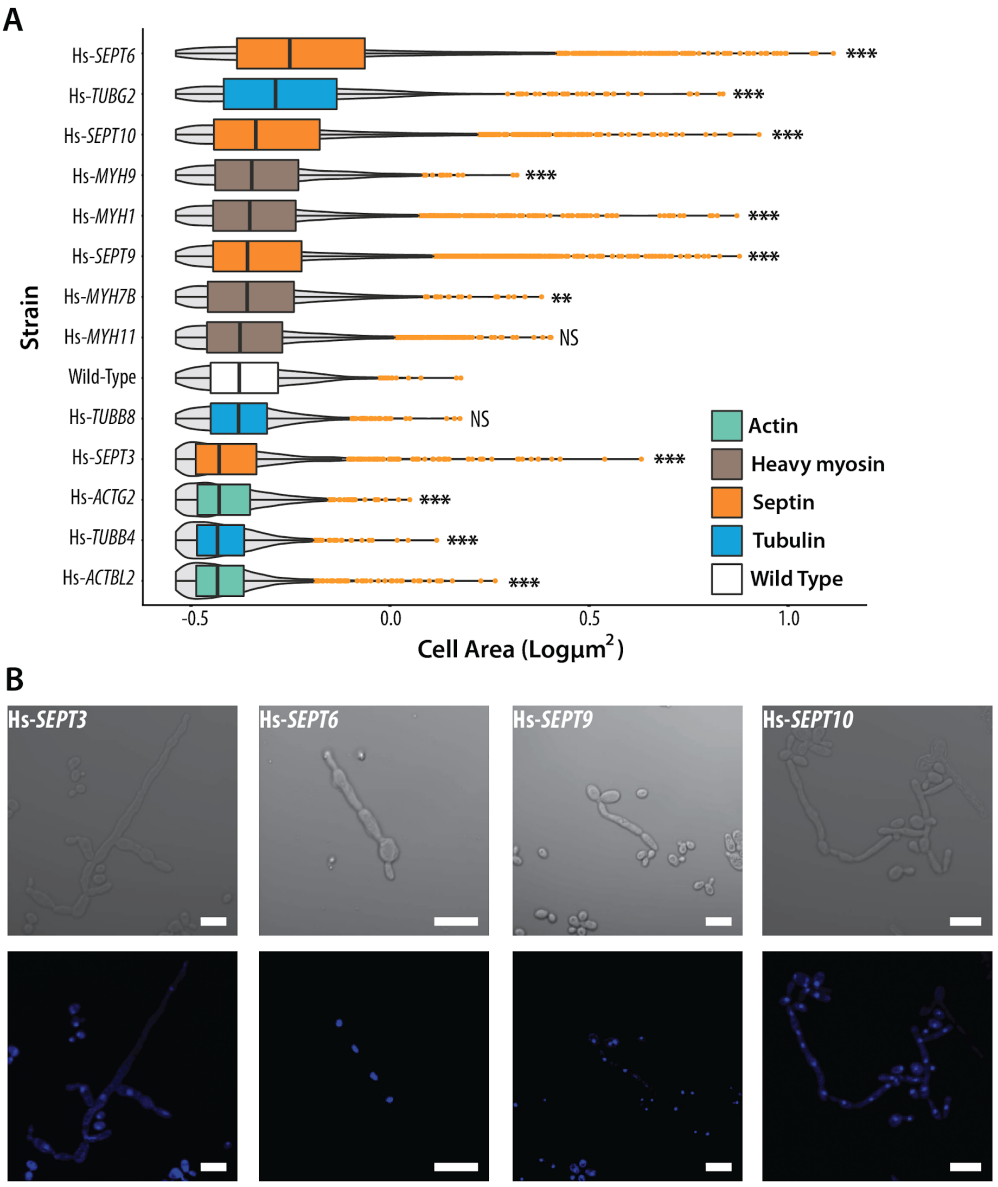


Figure 3.6.6. Yeast strains with humanized cytoskeletal components exhibit distinct cellular morphologies.

(A) Humanized strains show varying cell sizes. Cell areas of humanized strains (in square pixels) are plotted on the X-axis. Grey violins indicate cell size distributions. Orange dots

indicate outliers. Humanized actin strains show reduced cell size while myosins and tubulins (except ɣ-tubulin) largely remain unchanged. However, humanized septin strains show drastically elongated cellular morphologies. Significance comparisons with wild-type determined by standard t-test with ***, $p \leq 0.001$; **, $0.001 < p \leq 0.01$; *, $0.01 < p \leq 0.05$; NS, $p > 0.05$, not significant. Additional bright-field images are shown **Fig. 3.6.S8.** (B) Magnified bright-field and DAPI-stained images of humanized septin strains exhibit elongated morphologies and are multinucleated as a consequence of defective cytokinesis (Scale bars indicate 10 μm).

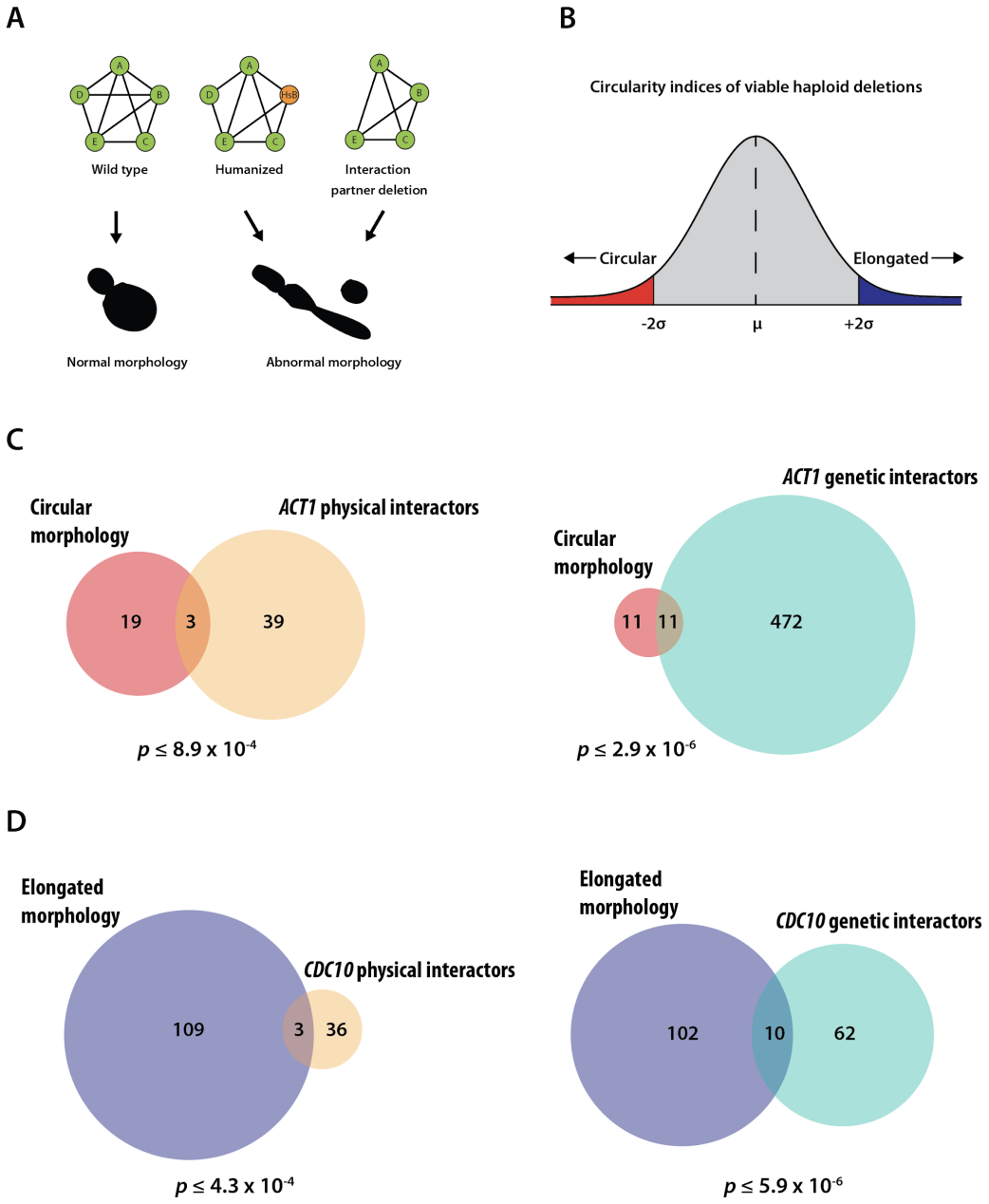


Figure 3.6.7. Humanized cytoskeletal yeast strains with abnormal cell morphologies phenocopy deletions of their corresponding yeast ortholog's interaction partners.

(A) Schematics of genetic interactions in the wild-type (Left), humanized (Middle), and interacting partner yeast deletion strains. Humanized yeast and yeast deletion strains are

shown to cause similar abnormal cell shapes. (B) Cell shape parameters calculated from SCMD and genetic interactors of the replaced yeast ortholog curated from SGD show significant overlap (*p*-values determined by hypergeometric test).

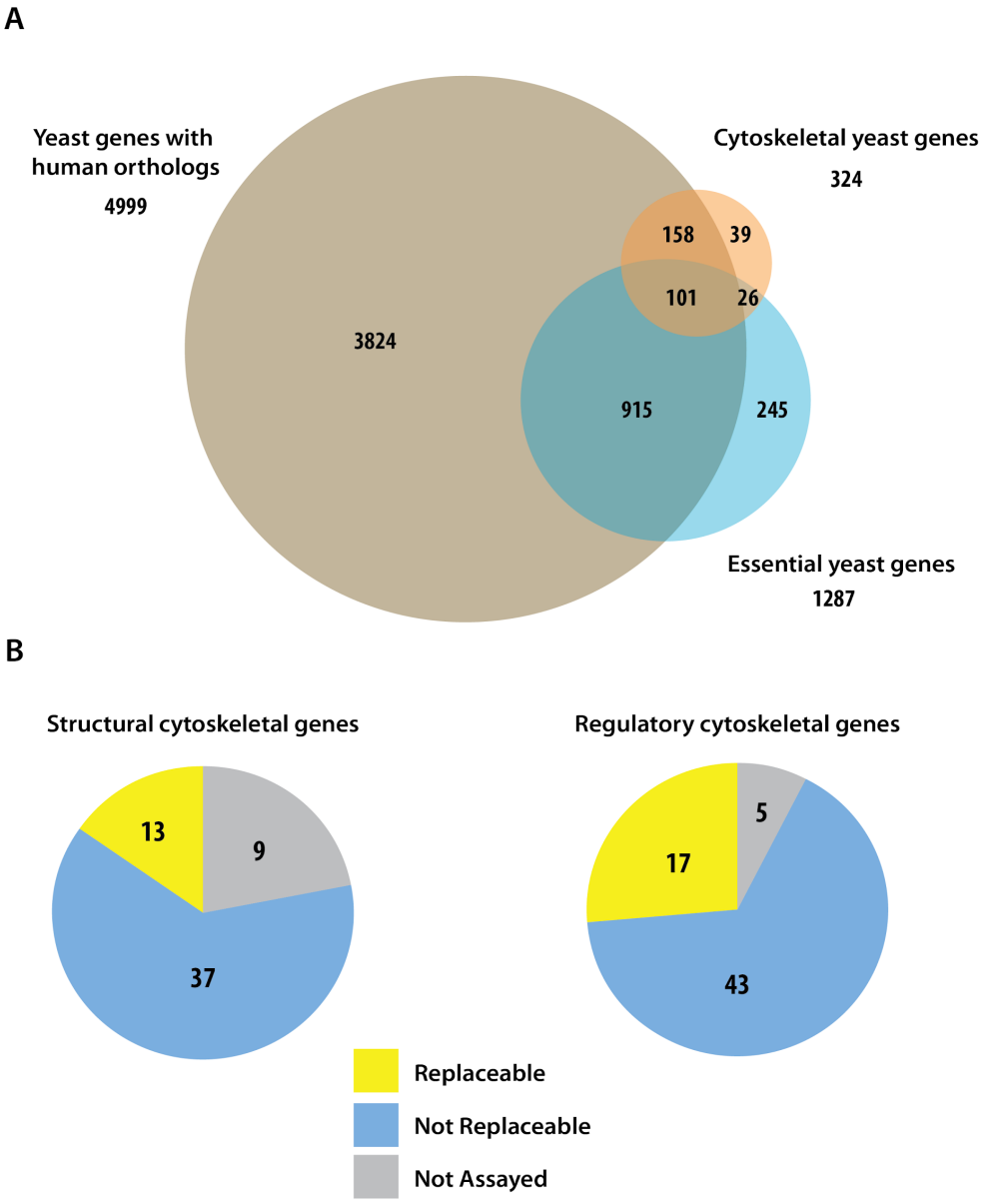


Figure 3.6.S1. Humans and yeast share a substantial number of orthologous genes.

(A) Global orthology relationships shared between human and yeast genes. Yeast genes with human orthologs are represented in brown, cytoskeleton, and essential yeast genes in

orange and blue, respectively. (B) Replaceability distribution of human genes across structural and regulatory elements of the eukaryotic cytoskeleton. Data for regulatory human cytoskeletal elements from Laurent *et al* (Laurent *et al.* 2019).

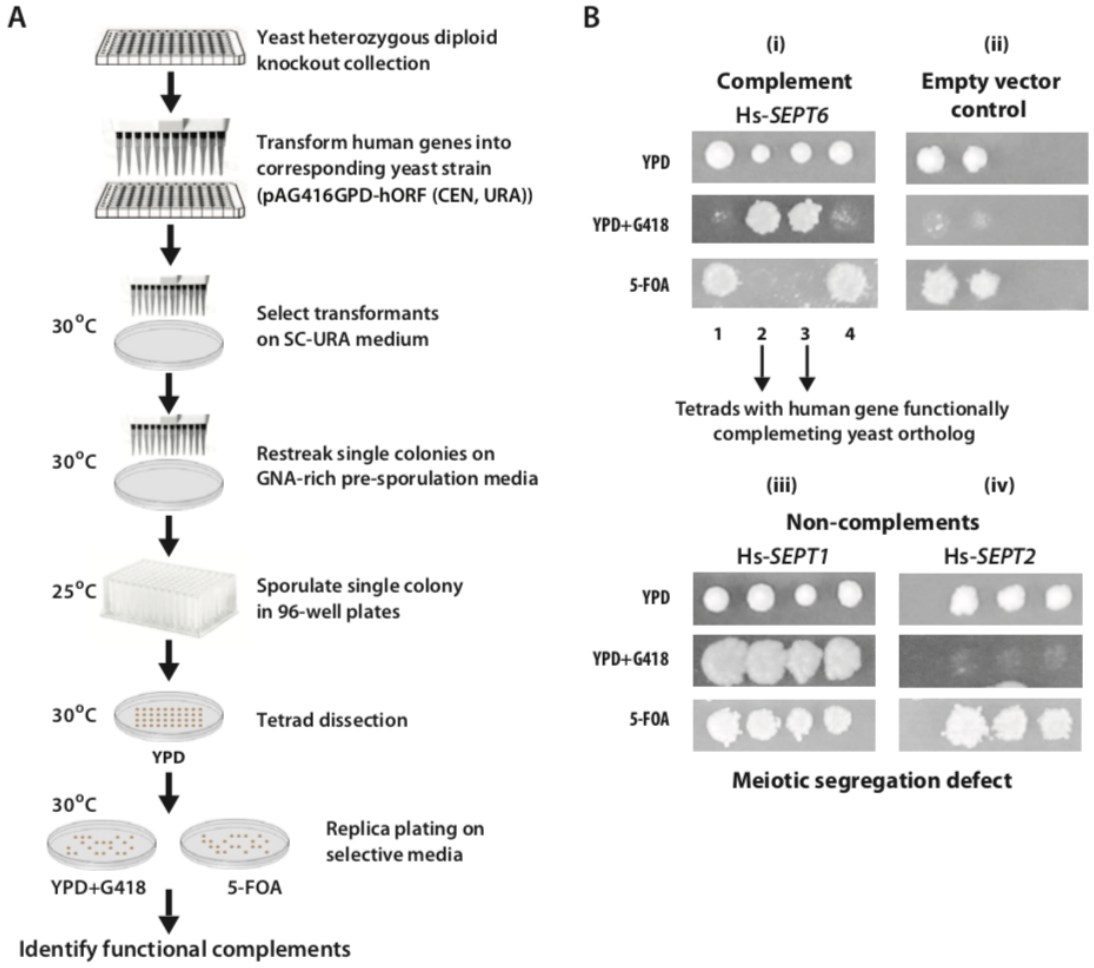


Figure 3.6.S2. Representative heterozygous diploid deletion allele complementation assay.

(A) Schematic outline of humanization assays performed in heterozygous diploid deletion yeast strains. (B) Complementation assay performed after sporulation and tetrad dissection shows examples of (i) replaceable, (ii) empty vector control, and (iii) non-complementing cases.

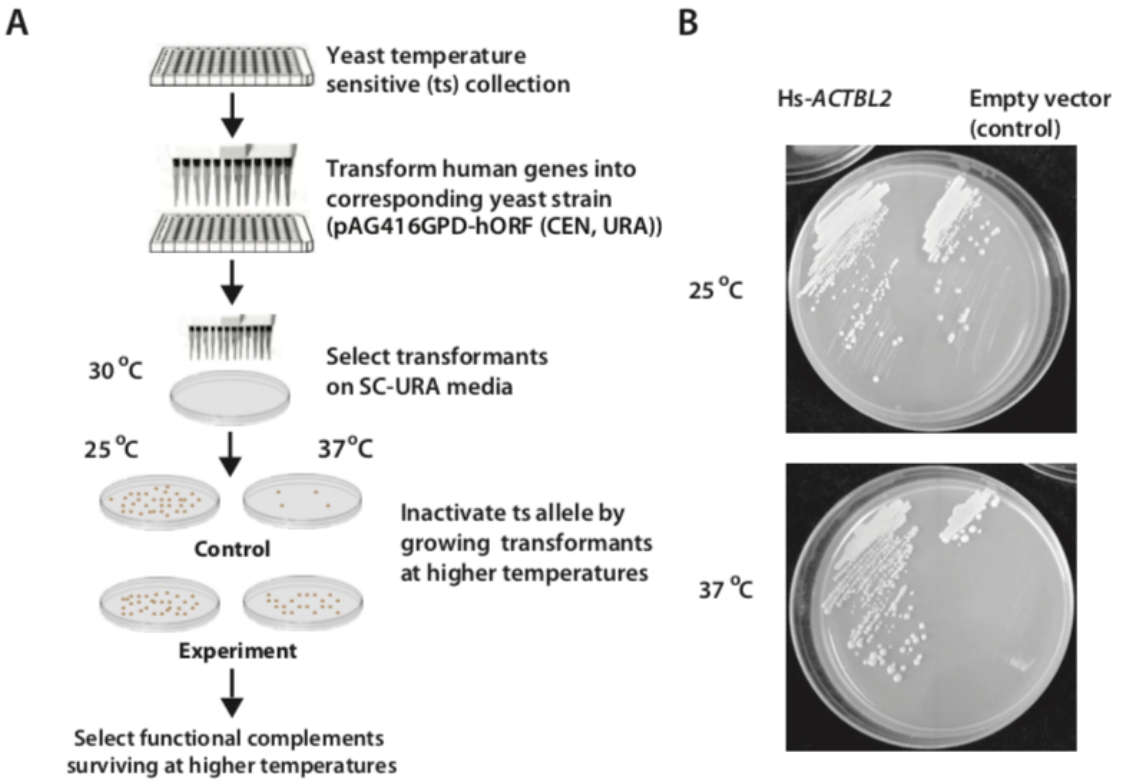


Figure 3.6.S3. Representative temperature-sensitive allele complementation assay.

(A) Schematic outline of the assays performed in temperature-sensitive haploid yeast strains. (B) Bottom plate showing that *Hs-ACTBL2* functionally replaces the temperature dependent lethal growth defect observed at 37ºC (left) as opposed to its empty vector control (right). Top plate indicates control experiments at 25ºC.

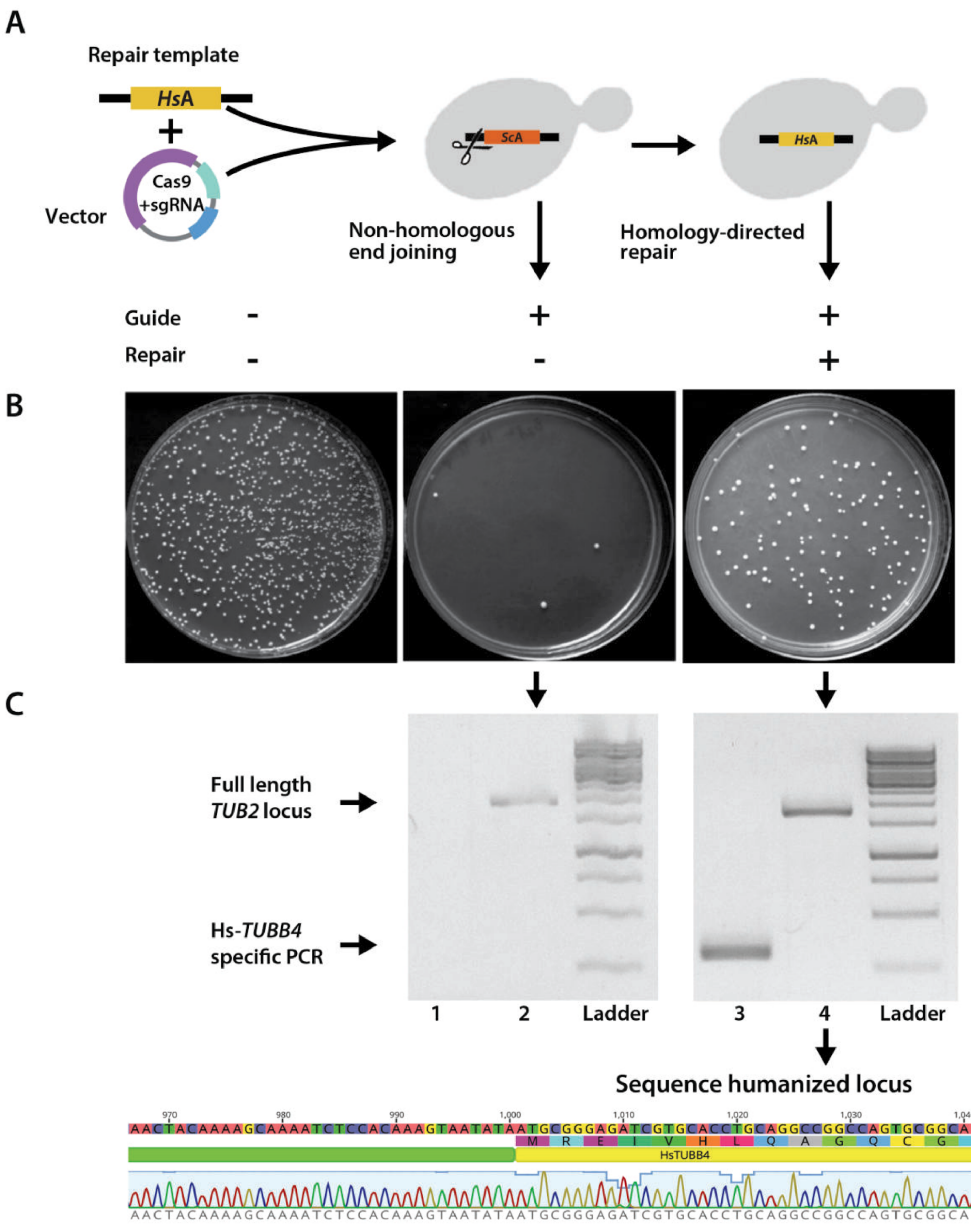


Figure 3.6.S4. Representative CRISPR-Cas9 complementation assay.

(A) Schematic outline of humanization of yeast genes at their native yeast loci using CRISPR-Cas9 based genome editing (B) Yeast cells are rescued from lethal double

stranded breaks caused by the Cas9-sgRNA complex (center plate) by supplying a human ortholog containing repair template (right plate) with flanking sequence homology to yeast locus of interest. The left plate is a negative control (carrying the same selectable marker) without Cas9 and sgRNA expressing transcription units, performed to estimate competent cell transformation efficiency. (C) Colony PCR (top) and Sanger sequencing (bottom) to confirm the integration of the human ortholog Hs-*TUBB4* (Lanes 1 & 3) into the corresponding yeast *TUB2* genomic locus (Lanes 2 & 4) using primers binding outside the region of homology provided for the repair of the double strand break.

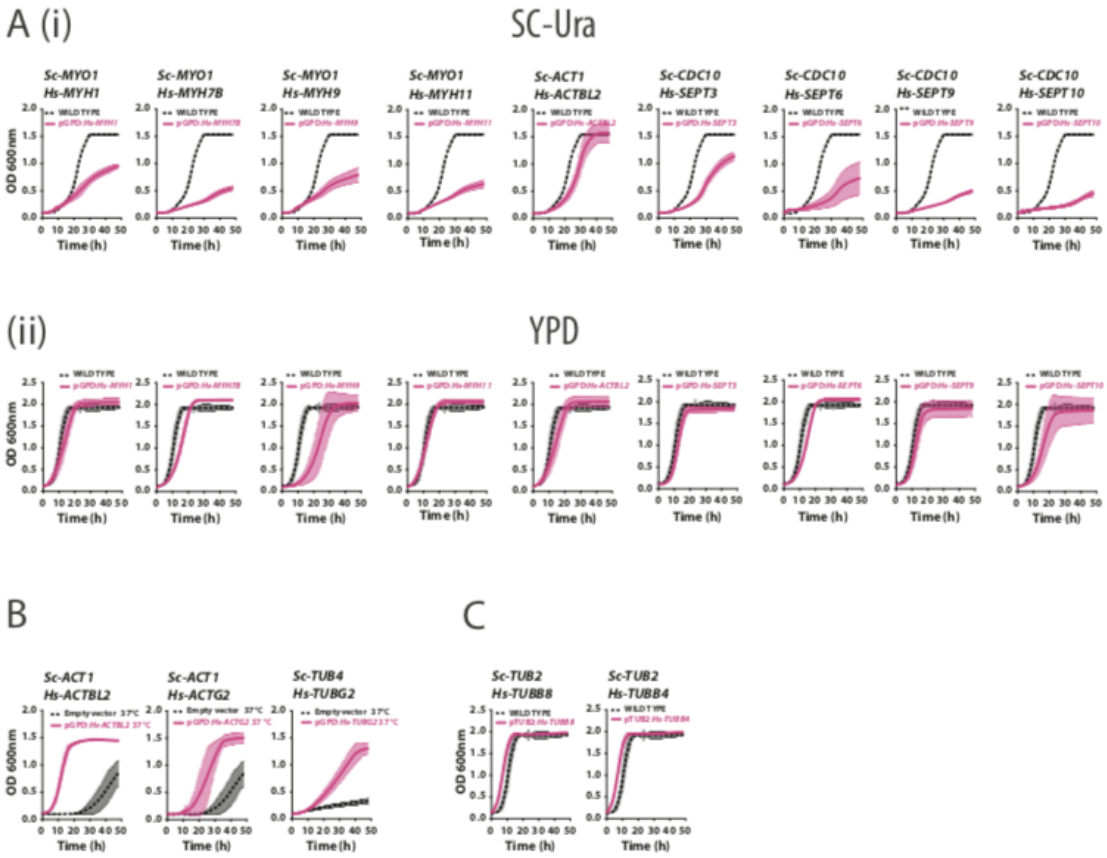


Figure 3.6.S5. Growth profiles of humanized strains.

In all, 89 of 109 possible human-yeast complementation pairs were tested, amounting to 50 of 62 human genes. We observe that 13 human genes complement the lethal loss of the corresponding yeast ortholog. Quantitative growth assays of humanized strains measuring absorbance at 600nm assayed *via* (A) Plasmid based complementation of (haploid) humanized yeast strains after sporulation, tetrad dissection and 5-FOA/G418 screening of heterozygous deletion diploids in (i) SC-Ura or (ii) YPD rich medium at 30ºC. We observe substantial growth rate differences across both media even though the human gene encoding expression vector is still retained in the rich medium. (B) Plasmid

based complementation of temperature-sensitive strains performed at 37ºC. (C) Growth profiles of haploid yeast strains at 30ºC after humanization of yeast genes at their native genomic loci *via* CRISPR-Cas9. Solid lines represent the mean of Wild-type (black) and humanized (magenta) and shaded boundaries demonstrate +/- standard deviation of 3 replicates each.

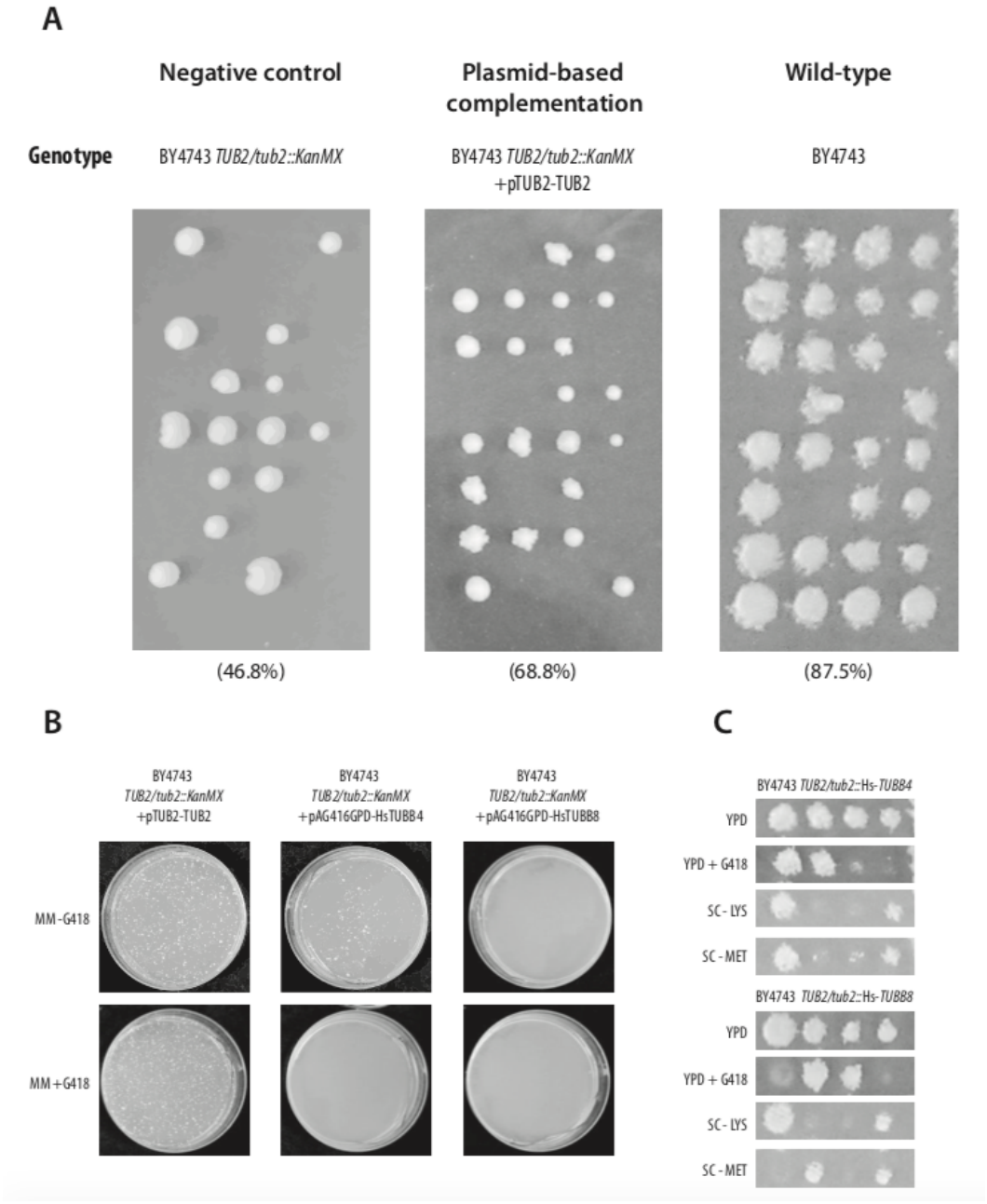


Figure 3.6.S6. Tetrad dissection and analysis of humanized β-tubulin strains.

(A) Sporulation and tetrad dissection of BY4743 *TUB2*/*tub2::KanMX* heterozygous knockout strain serves as a negative control in this experiment (left panel) and the yeast

strain harboring plasmid borne copy of the yeast *TUB2* gene (center panel). Wild-type homozygous diploid (right panel) serves as a positive control. Plasmid borne *TUB2* rescues the loss-of-function of the yeast gene at the native locus. (B) Plasmid complementation assays of yeast *TUB2* by *TUB2* (left), Hs-*TUBB4* (center), Hs-*TUBB8* (right) on SC-Arg-His-Leu+Can-Ura +/- G418. As in the case of *TUB2* (left), a complementation consists of spores able to proliferate both in the presence and absence of G418. In the case of Hs-*TUBB4* (center) selecting for spores harboring the *TUB2* null allele (+G418 condition) is lethal as the human gene on the plasmid does not complement. However, in the case of Hs-*TUBB8* (right) the expression of the gene tends to be toxic and suggestive of defective sporulation since no spores are able to grow even in the -G418 condition. (C) Analysis of segregation of genetic markers in humanized tubulin strains. Humanized strains heterozygous diploid for Hs-*TUBB4* (left panel) and Hs-*TUBB8* (right panel) were sporulated, tetrad-dissected, and selected on MET^- (harbored by BY4741) and LYS (harbored by BY4742) media to check individual loci for 2:2 independent segregation. Since we performed tubulin humanizations in a BY4741 *tub3*::KanMX background, we also scored G418 resistance, which also segregated 2:2 manner.

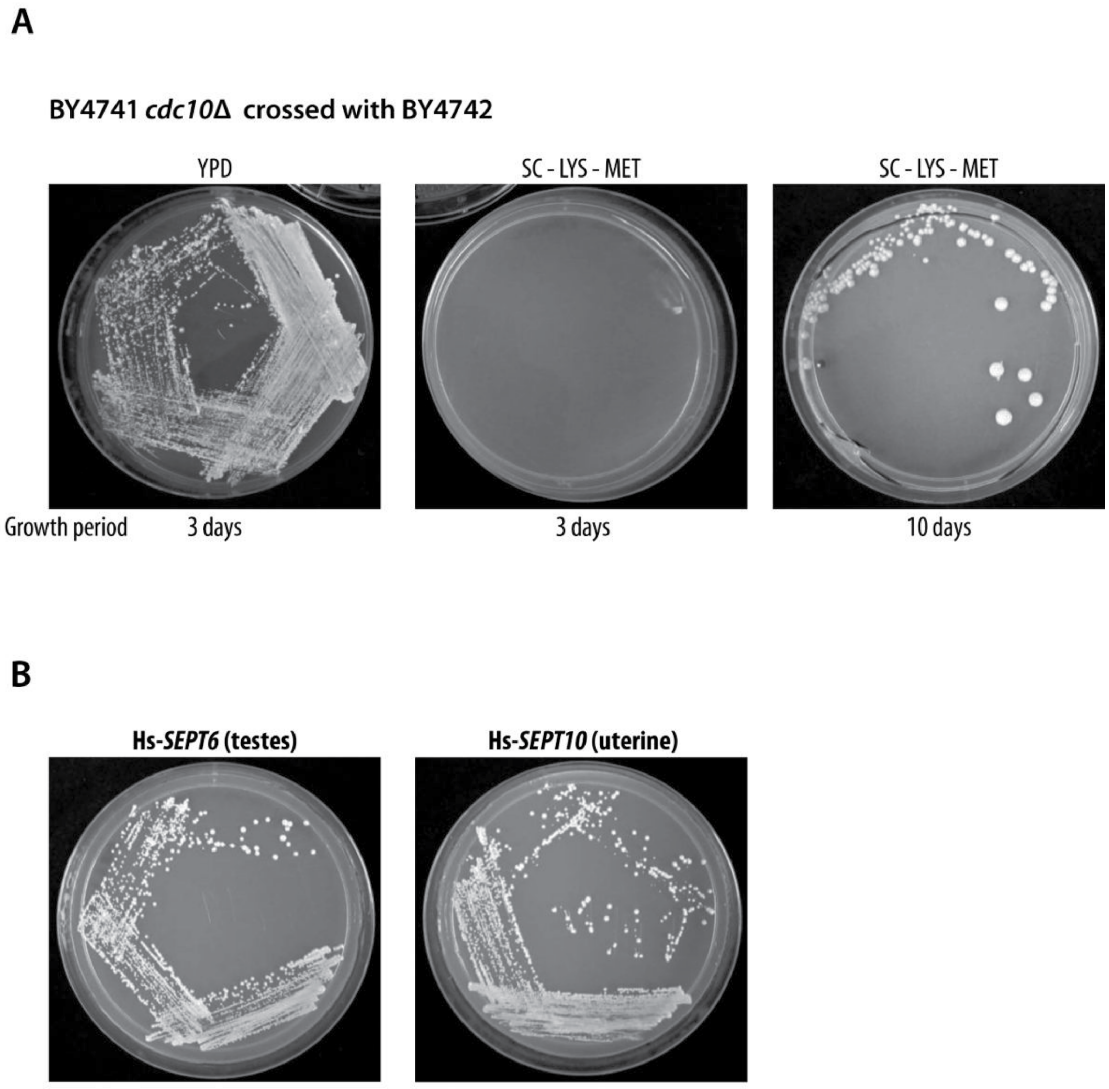


Figure 3.6.S7. Humanized septin mating assays.

(A) Mating assays performed with BY4741 *cdc10Δ* strains show severe mating defects resulting from deletion of the yeast septin *CDC10*. BY4741 *cdc10Δ* strains show severe mating defects taking ~10 days to produce diploids, while complementation with human septins rescues this mating defect. (See **Fig. 3.6.5B**) (B) Human septin isoforms along

with other septins, Hs-*SEPT6* and Hs-*SEPT10* can individually rescue the mating defect caused by the deletion of the yeast ortholog, *CDC10*. Plate images depict diploids selected on SC-Lys-Met medium.

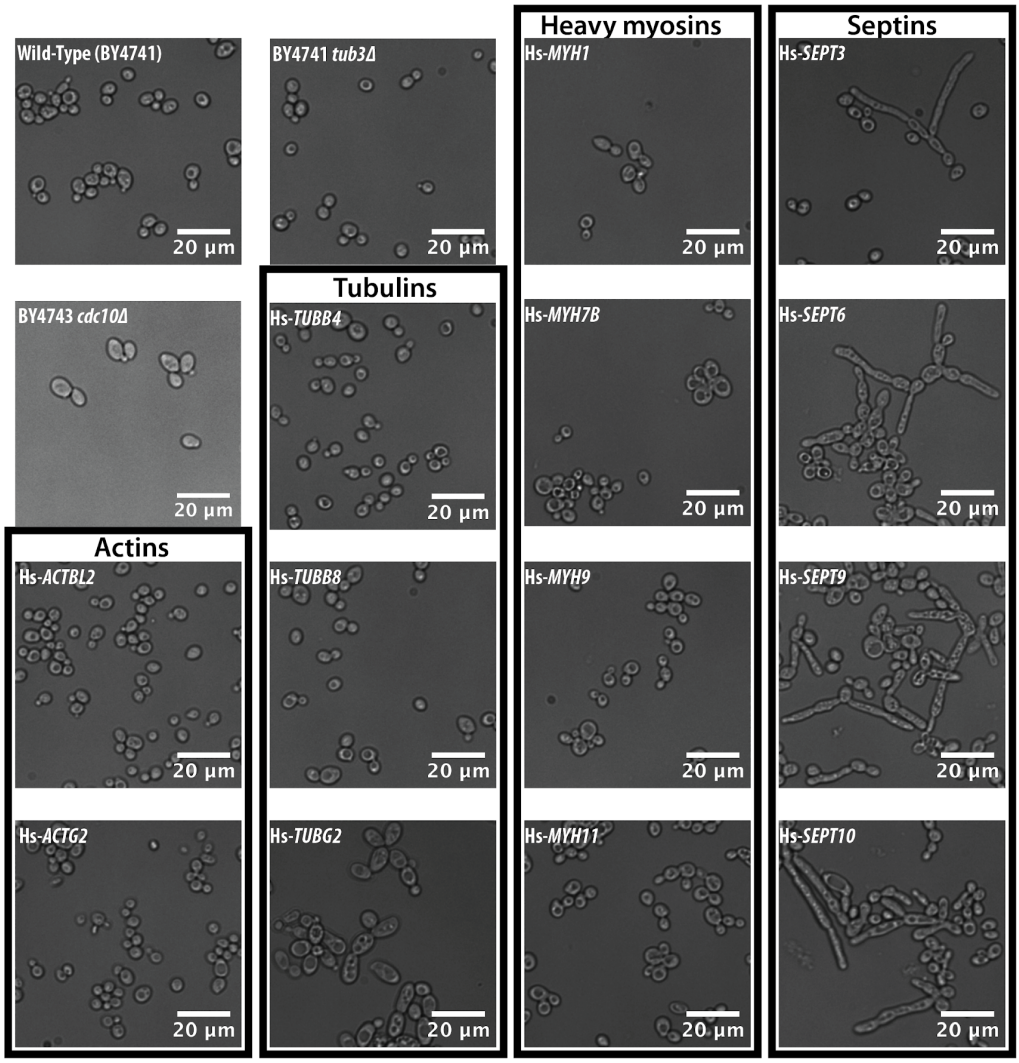


Figure 3.6.S8. Yeast strains with replaceable human cytoskeletal genes exhibit abnormal cell morphologies.

Bright-field images of humanized yeast strains sorted by family.

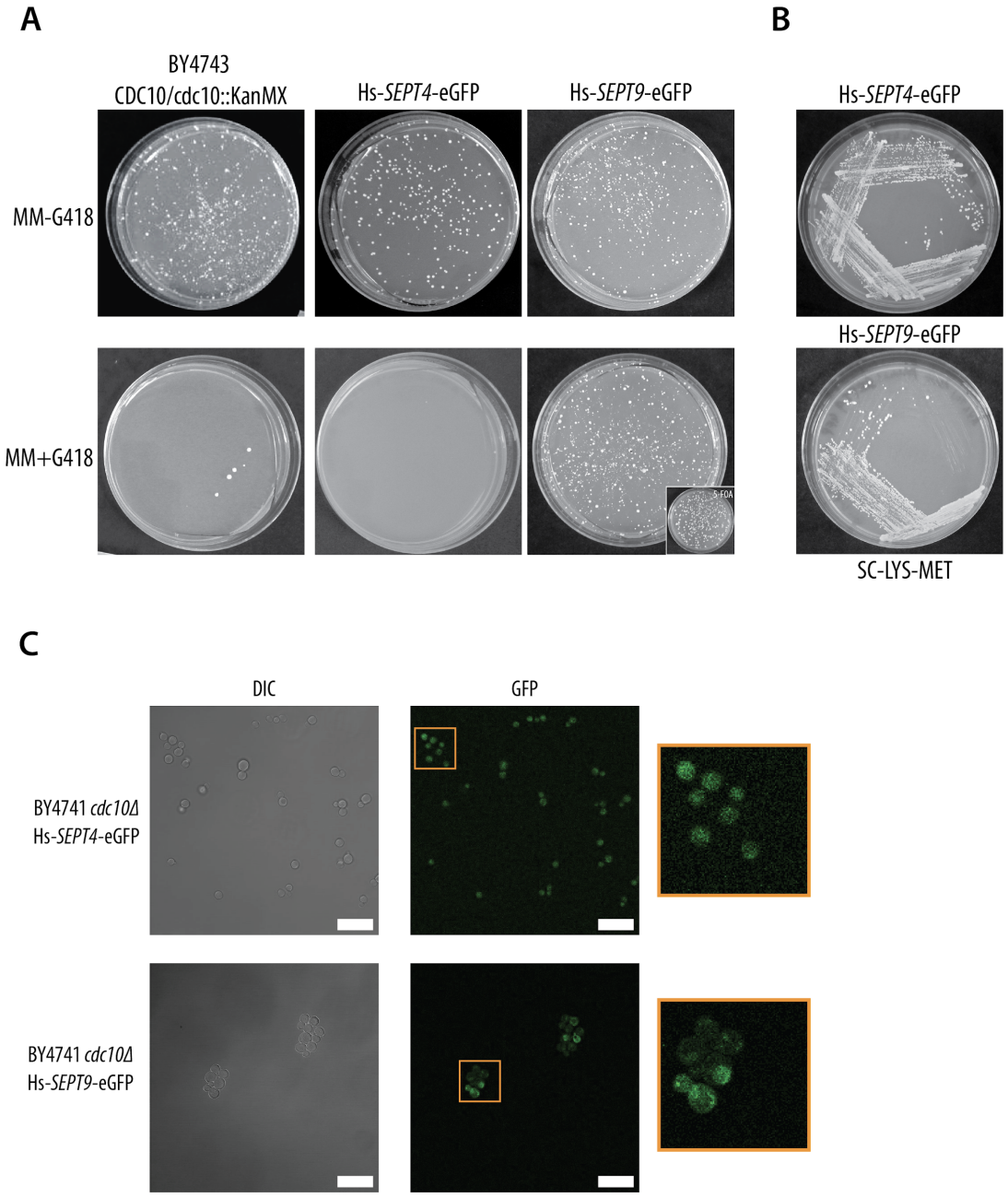


Figure 3.6.S9. GFP tagged septins also functionally complement *CDC10*.

(A) Complementation assays of GFP-tagged human septins. Plasmid complementation assays of *CDC10* by empty vector control (left), Hs-*SEPT4-eGFP* (center) and Hs-

SEPT9-eGFP (right) on SC-Arg-His-Leu+Can-Ura-G418 (top) and +G418 (bottom). A complementation consists of spores able to proliferate both in the presence and absence of G418. In the case of the empty vector control (left) and Hs-*SEPT4-eGFP* (center) selecting for spores harboring the *CDC10* null allele (+G418 condition) is lethal as both the empty vector and human gene on the plasmid fail to complement. However in the case of Hs-*SEPT9-eGFP* (right), expression of the human gene rescues the sporulation defect since colonies are able to proliferate on both + and -G418 conditions. The inset shows the replica-plating of the +G418 Hs-*SEPT9*-eGFP colonies also growing on 5-FOA indicating segregation defects induced by human gene expression. (B) GFP tagged versions of both Hs-*SEPT4* and Hs-*SEPT9* can rescue the mating defect observed in *cdc10Δ* strains similar to their untagged versions. But Hs-*SEPT4-eGFP* similarly fails to rescue the sporulation defect observed with its untagged version (C) GFP-tagged human septins Hs-*SEPT4* and Hs-*SEPT9* do not localize to the bud neck as seen in the case of yeast *CDC10*. Left panels indicate the DIC (brightfield) channel and right panels indicate GFP channel with 20 micron scale bar width. Orange boxes depict zooms of the respective cells in their respective fluorescence channels.

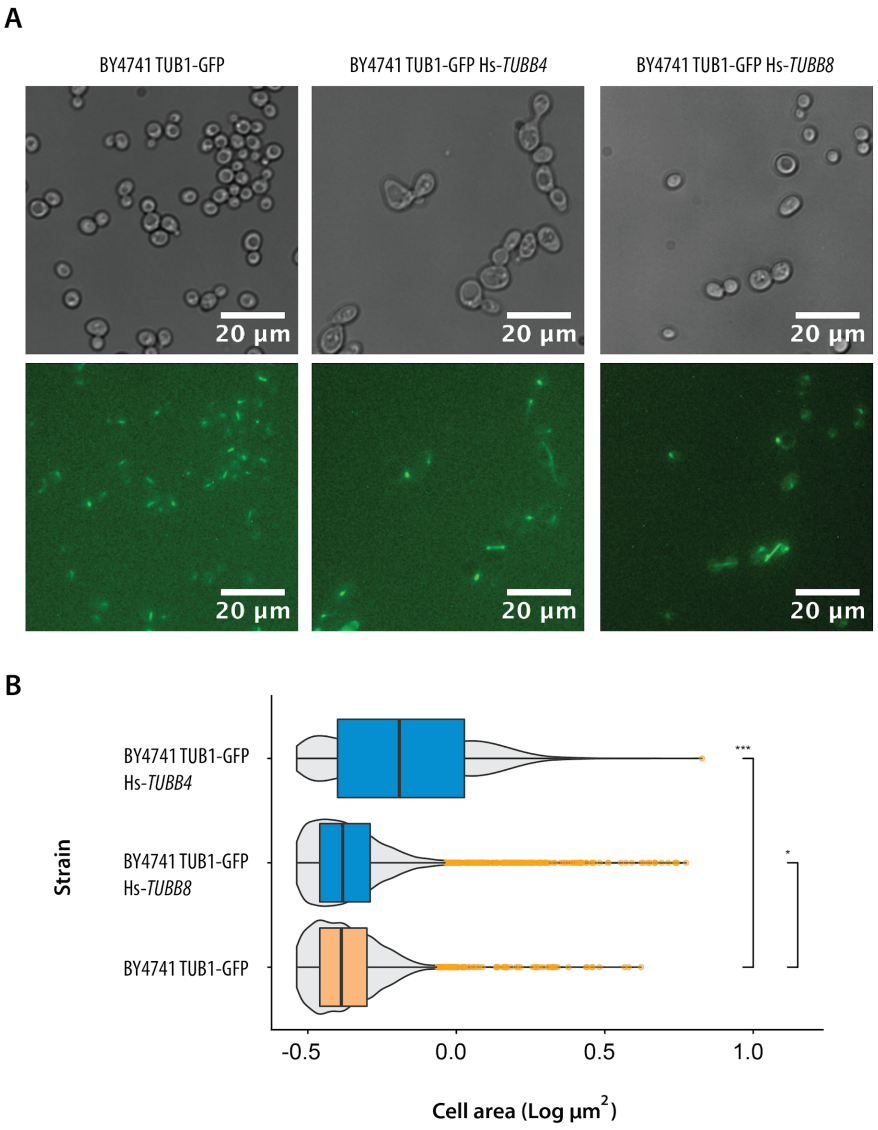


Figure 3.6.S10. Human β-tubulins are successfully incorporated into the microtubule structure.

(A) GFP-fluorescence imaging of the wild-type haploid yeast strain with yeast α-tubulin *TUB1* tagged with GFP (left panel) and harboring human β-tubulins, Hs-*TUBB4* (center panel) and Hs-*TUBB8* (right panel). (B) Cell size distributions of the strains imaged in

(A) reveal an obvious cell size increase when replacing Hs-*TUBB4* in a *TUB1-GFP* background. Grey violins indicate size distributions. Significance comparisons were calculated using standard t-test and indicated as in Fig. 3.6.6.

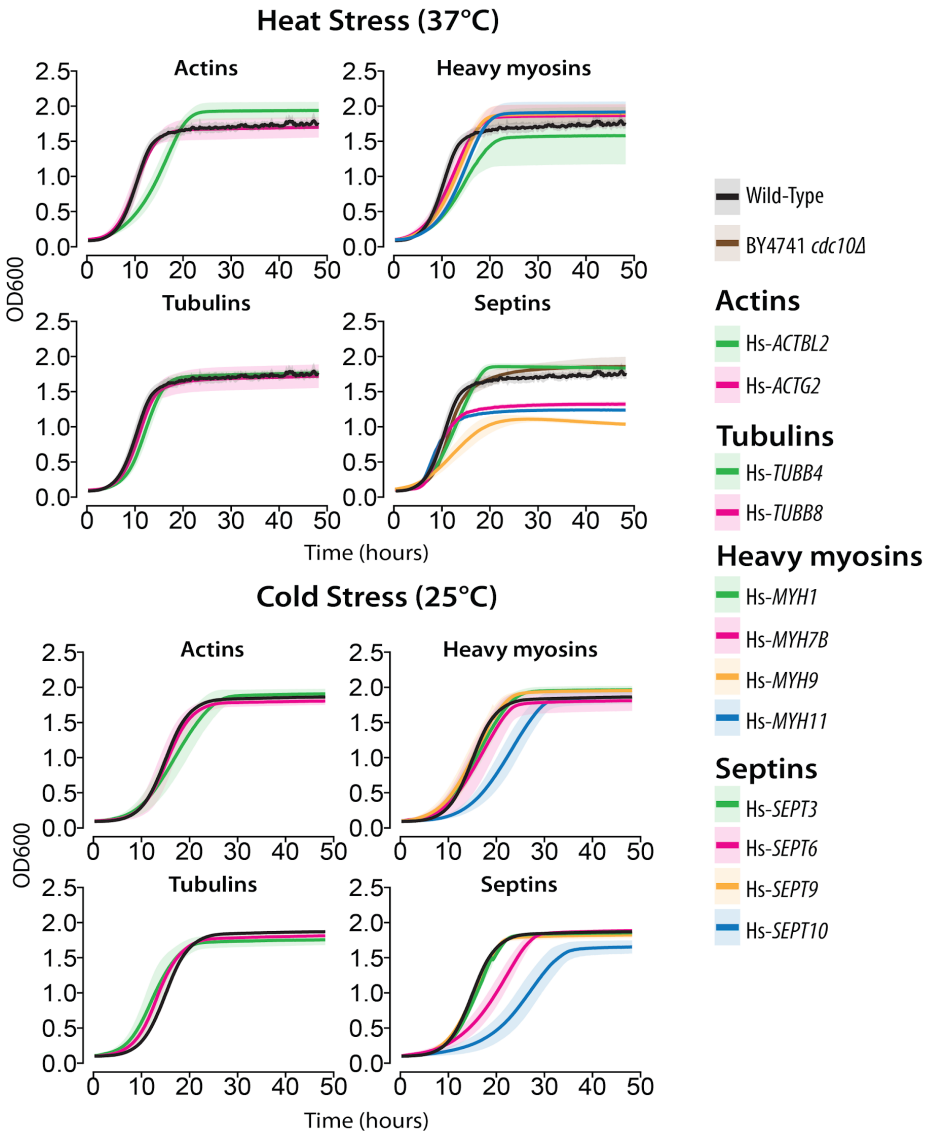


Figure 3.6.S11. Humanized cytoskeletal strains demonstrate temperature dependent fitness defects.

Quantitative growth assays on humanized strains subjected heat (37°C) and cold (25°C) stress. The mean (solid line) +/- standard deviation (shaded boundary) is plotted for 3 replicates.

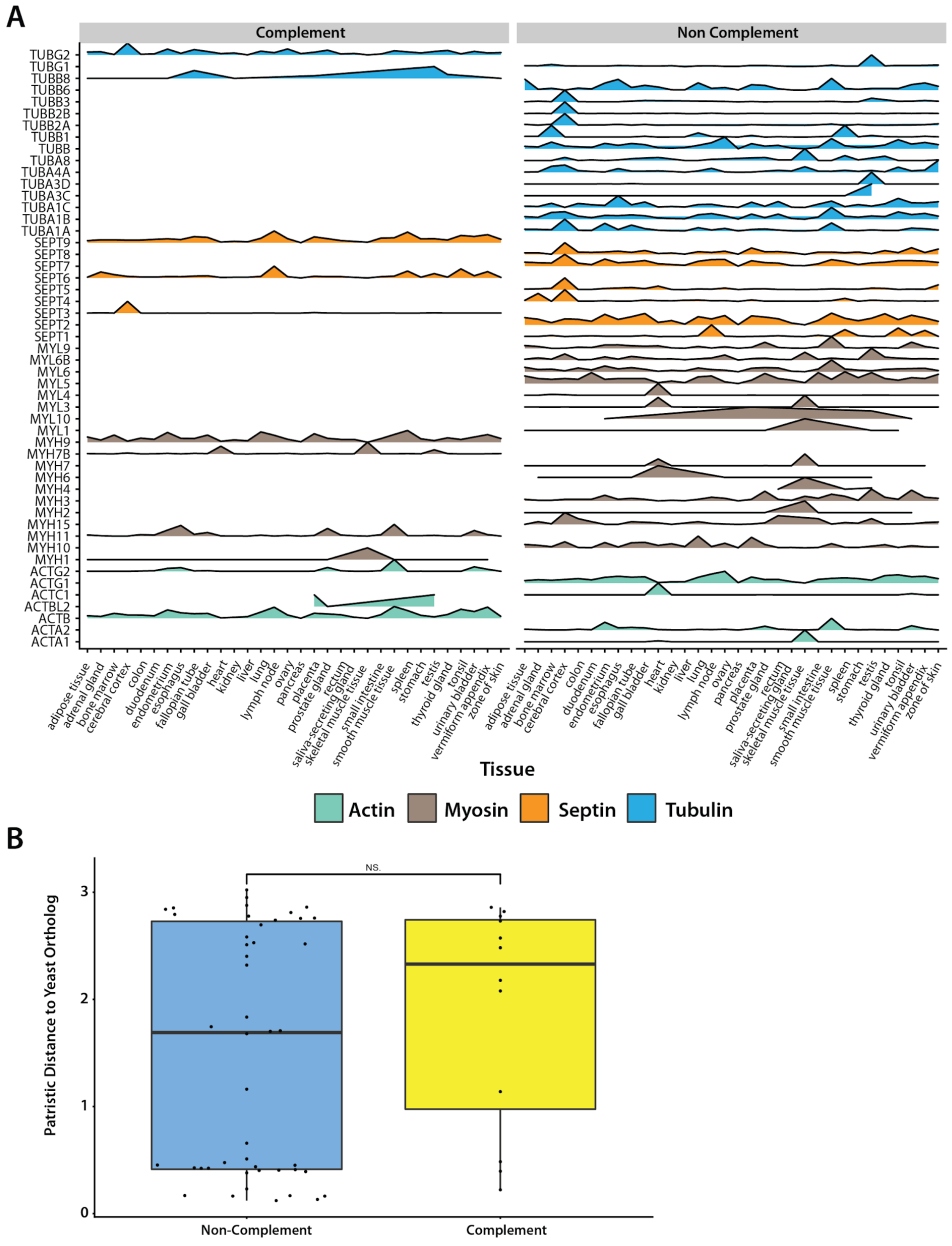


Figure 3.6.S12. Tissue expression patterns and sequence divergence of human cytoskeletal genes do not explain replaceability.

(A) Tissue expression data was curated from Human Protein Atlas for available human cytoskeletal genes. X axis represents tissue type, Y axis represents cytoskeletal gene.

Ridgeplots are binned by complementation status. (B) Patristic distance between human and yeast ortholog pairs does not predict replaceability. Boxplot comparing patristic distances of a human-yeast ortholog pair in cytoskeletal families. Patristic distance computed from multiple sequence alignment from MAFFT.

3.7. TABLES

Strain	Growth rate in YPD (/hrs)	Growth rate in URA (/hrs)	Doubling time in YPD (hrs)	Doubling time in URA (hrs)
Hs-TUBB4	0.5387127	NA	1.286673	NA
Wild-type	0.5349355	0.31743107	1.295758	2.183615
Hs-TUBB8	0.5306934	NA	1.306116	NA
Hs-SEPT3	0.5014354	0.17186381	1.382326	4.033119
Hs-SEPT9	0.4563149	0.0981111	1.519011	7.064921
Hs-SEPT6	0.3638143	0.11498014	1.905223	6.028408
Hs-MYH11	0.342426	0.12376502	2.024225	5.60051
Hs-MYH7B	0.3265971	0.10264045	2.122331	6.753158
Hs-SEPT10	0.324499	0.05488795	2.136053	12.628404
Hs-MYH1	0.3105858	0.13957244	2.231741	4.966218
Hs-ACTBL2	0.2970452	0.22531536	2.333473	3.076342
Hs-MYH9	0.2470456	0.13523041	2.805746	5.125675
Hs-ACTG2	0.2295563	NA	3.019508	NA
Hs-TUBG2	0.1349441	NA	5.13655	NA

Table 3.7.1. Growth characteristics of humanized yeast strains.

Chapter 4: Antifungal benzimidazoles disrupt vasculature by targeting one of nine β-tubulins[3]

This final chapter encapsulates what is arguably the most exciting and favorite years of my graduate work because it tied together multiple different threads in the lab and synthesized in an exciting discovery. The origin of this project dates back 8-10 years before I started my Ph.D. As with most projects in the Marcotte lab, this too was motivated by understanding how orthologous modules between species functionally diverge over time. Over the past several years, the lab established "phenologs" a concept that defines orthologous phenotypes between organisms by systematically overlapping sets of orthologous genes associated with each organism-specific phenotype (McGary *et al.* 2010). In other words, two phenotypes (between species) are said to be phenologs if they are linked by a set of orthologous genes. Most interestingly, this approach supplies links between conserved gene modules and how they likely work in organism-specific contexts. Besides mapping these seemingly non-obvious phenotypic associations between pairs of organisms, this novel approach allows for the identification of new candidate genes that are likely implicated in disease. A couple of notable examples include a *Saccharomyces cerevisiae* (Baker's yeast) model for angiogenesis and an *Arabidopsis thaliana* (thale cress plant) model for Waardenburg syndrome. Obviously, yeast lack blood vessels (and angiogenesis) and *Arabidopsis thaliana* lack ears (to suffer from congenital hearing loss) but such orthologous gene sets involved in their corresponding

[3]This chapter was published as Garge, R.K., Cha, H.J., Lee, C., Gollihar, J.D., Kachroo, A.H., Wallingford, J.B., and Marcotte, E.M. (2020). Antifungal benzimidazoles disrupt vasculature by targeting one of nine β-tubulins. *BioRxiv*. This work was co-led by Hye Ji Cha and myself. My contributions were performing the yeast assays, bioinformatic analyses and writing of the manuscript.

phenotypes (i.e., cell wall development in yeast and gravitropism) provide unexplored routes to study angiogenesis and Waardenburg syndrome respectively. Extending this principle further, combining chemical screens with these phenotypic relationships can be fruitful in discovering and developing therapies against human diseases. In 2014, by leveraging the yeast model of angiogenesis, Hye Ji Cha found that the widely used FDA-approved antifungal Thiabendazole inhibited angiogenesis functioning as a vascular disrupting agent (Cha *et al.* 2012).

While this phenolog-driven approach resulted in an exciting discovery, Thiabendazole's molecular target and mechanism of vascular action still remained unknown. This chapter describes how Thiabendazole's target was uncovered using a combination of *in silico* homology modeling with docking, tissue culture, model organisms, humanized yeast, and mining chemical properties of benzimidazoles with epidemiological signatures of pesticide resistance. This chapter is the synthesis of multiple research threads that resulted in an even more exciting discovery of 5 more compounds (similar to Thiabendazole) all functioning as vascular disrupting agents. Most interestingly, this group included two human intestinal antihelminthics prescribed by the World Health Organization making a case to reevaluate their use in young children and pregnant women. Finally, this chapter is a poster child to the power that model organisms, systems biology, and evolution-guided approaches wield when used together.

4.1. ABSTRACT

Thiabendazole (TBZ) is an FDA-approved benzimidazole widely used for its antifungal and antihelminthic properties. We showed previously that TBZ is also a potent vascular disrupting agent and inhibits angiogenesis at the tissue level by dissociating vascular endothelial cells in newly formed blood vessels. Here, we uncover TBZ's molecular target and mechanism of action. Using human cell culture, molecular modeling, and humanized yeast, we find that TBZ selectively targets only 1 of 9 human β-tubulin isotypes (TUBB8) to specifically disrupt endothelial cell microtubules. By leveraging epidemiological pesticide resistance data and mining chemical features of commercially used benzimidazoles, we discover that a broader class of benzimidazole compounds, in extensive use for 50 years, also potently disrupt immature blood vessels and inhibit angiogenesis. Thus, besides identifying the molecular mechanism of benzimidazole-mediated vascular disruption, this study presents evidence relevant to the widespread use of these compounds while offering potential new clinical applications.

4.2. INTRODUCTION

The vascular system is built by the combination of *de novo* formation of blood vessels by vasculogenesis and the sprouting of new vessels from existing vessels via angiogenesis (Carmeliet 2005; Herbert and Stainier 2011). Imbalances in angiogenesis underlie a variety of physiological and pathological defects, including ischemic, inflammatory, and immune disorders (Carmeliet 2005; Folkman 2007; Kerbel 2008).

Indeed, angiogenesis is central to tumor malignancy and cancer progression, as new blood vessels must be established to supply oxygen and nutrients to the growing tumor. Accordingly, inhibition of angiogenesis is now a well-recognized therapeutic avenue (Folkman 2004, 2007; Carmeliet 2005; Kerbel 2008; Herbert and Stainier 2011). Defined angiogenesis inhibitors such as Avastin (FDA approved since 2004) are now in wide use in the clinic and, over the past 30 years, several dozen drugs have been approved or entered clinical trials as angiogenesis inhibitors (O'Reilly *et al.* 1994, 1997; Folkman 2004; Nyberg *et al.* 2005; El-Kenawi and El-Remessy 2013).

In recent years, a new class of anti-vascular drugs, termed vascular disrupting agents (VDAs), have gained attention as potential alternative therapeutics operating by distinct mechanisms (Lippert 2007; Hinnen and Eskens 2007; Heath and Bicknell 2009; Mason *et al.* 2011). Unlike angiogenesis inhibitors which selectively prevent the formation of new blood vessels, VDAs function by dismantling existing vasculature, making them potentially effective for therapies beyond cancer, for example in the treatment or control of macular degeneration and diabetic retinopathies (Ibrahim *et al.* 2013; Nowak-Sliwinska *et al.* 2013). While several VDAs have shown therapeutic potential, none have yet been approved, with several candidates still in clinical trials (Tozer *et al.* 2005; Cai 2007; Hinnen and Eskens 2007).

Given the lengthy approval process, the failure of many drugs to succeed in clinical trials, and the high costs involved with developing new compounds, drug repurposing offers an attractive alternative for developing new therapies more quickly. We recently developed strategies to exploit data from diverse model organisms to

identify both deeply conserved genetic networks as well as small molecules that may manipulate them (McGary *et al.* 2010; Cha *et al.* 2012; Woods *et al.* 2013). This effort identified thiabendazole (TBZ) as both a novel angiogenesis inhibitor and VDA (Cha *et al.* 2012).

TBZ is one of a large class of biologically active benzimidazole compounds that are widely used commercially or clinically, with applications ranging from photographic emulsions and circuit board manufacturing, to serving as one of the most common heterocyclic ring systems used for small molecule drugs (Taylor *et al.* 2014). The FDA approved TBZ in 1967 for human use for treating systemic fungal and helminthic infections, but it is more widely used in veterinary settings and in agricultural pesticides and preservatives. However, we found that TBZ also possesses potent vascular disrupting ability, demonstrated *in vitro* in human cell culture and *in vivo* in mice and frogs, including for retarding tumor growth and reducing intratumoral vessel density in preclinical murine xenograft models (Cha *et al.* 2012).

Several other VDAs have been reported to collapse the vasculature by inhibiting microtubule polymerization dynamics *via* binding β-tubulin (Tozer *et al.* 2005; Hinnen and Eskens 2007). Indeed, though the basis for TBZ's vascular disrupting action is unknown, it is proposed that TBZ's fungicidal action is mediated *via* disrupting fungal microtubule assembly and dynamics (Davidse and Flach 1978; Skuce *et al.* 2010). In particular, mutations in β-tubulin have been frequently found to confer resistance to TBZ in parasitic/invasive fungal and nematode species(Davidse and Flach 1978; Driscoll *et al.* 1989; Lubega and Prichard 1990; Lacey and Gill 1994; Skuce *et al.* 2010; Aguayo-Ortiz

et al. 2013; Taylor *et al.* 2014; Vela-Corcía *et al.* 2018; Hahnel *et al.* 2018). However, in humans, TBZ does not generally disrupt cell growth, and even in human umbilical vein endothelial (HUVEC) cells , while it somewhat reduced tubulin protein abundance it did not elicit gross defects in the microtubule cytoskeleton (Cha *et al.* 2012). At angiogenesis-inhibiting doses, the overall development of TBZ treated animals is normal (Cha *et al.* 2012), consistent with TBZ's safety record in humans and veterinary settings (EPA 2002). Therore, we hypothesized that only certain types of human cells, such as subsets of endothelial cells involved in forming the vasculature, might be uniquely susceptible to TBZ.

Here, we experimentally determined TBZ's specific molecular target and cellular mechanism of vascular disrupting activity. We find that TBZ disrupts microtubule growth, with increased potency in endothelial cells. Using predictive molecular modeling, human cell culture, and humanized yeast, we find TBZ predominantly targets only one of nine human β-tubulins, suggesting an explanation for its cell-type specificity. Finally, based on epidemiological data mining and chemical structures, we discovered that a larger family of benzimidazoles—in clinical and commercial use for >50 years—all act as VDAs, disrupting the vasculature in a vertebrate animal model. These newly discovered VDAs include two World Health Organization (WHO) antihelminthics (albendazole and mebendazole) administered for the treatment of human intestinal infections, one broad-spectrum antifungal/antihelminthic (fenbendazole) used to treat farm animal infections, and two banned pesticides (benomyl and carbendazim) used to prevent wild fungal and nematode mediated crop destruction. Knowledge of their

vascular disrupting activities should thus inform their use in at-risk individuals (such as during pregnancy) and opens new clinical applications for these compounds.

4.3. MATERIALS AND METHODS

4.3.1. Multiple sequence alignment

Human gene sequences were downloaded from the Uniprot database. The multiple sequence alignment for *S. pombe*, *S. cerevisiae*, and 9 human β-tubulin genes was constructed using MAFFT v7 (Katoh and Standley 2013) and visualized in Geneious v10 (https://www.geneious.com).

4.3.2. Molecular modeling of β-tubulins

Homology models of human and fungal β-tubulins were constructed using as a reference structure the previously determined *Ovis aries* β-tubulin crystal structures (PDB: 3UT5 and 3N2G) (Barbier *et al.* 2010; Ranaivoson *et al.* 2012). The template was prepared using the Molecular Operating Environment (MOE.09.2014) software package from Chemical Computing Group. The structure was inspected for anomalies and protonated/charged with the Protonate3D subroutine (310K, pH 7.4, 0.1 M salt) (Labute 2009). The protonated structure was then lightly tethered to reduce significant deviation from the empirically determined coordinates and minimized using the Amber10:EHT forcefield with R-field treatment of electrostatics to an RMS gradient of 0.1 kcal mol^{-1} $Å^{-1}$. Homology models of the wild-type fungal β-tubulin were prepared by creating 25 main chain models with 25 sidechain samples at 298K (625 total) within MOE. Intermediates

were refined to an RMS gradient of 1 kcal mol^{-1} $Å^{-1}$, scored with the GB/VI methodology, minimized again to an RMS gradient of 0.5 kcal mol^{-1} $Å^{-1}$ and protonated. The final model for each variant was further refined by placing the protein within a 6 Å water sphere and minimizing the solvent enclosed structure to an RMS gradient of 0.001 kcal mol^{-1} $Å^{-1}$. Models were evaluated by calculating Phi-Psi angles and superimposed against the reference structure. Homology models for each human β-tubulin were prepared similarly, based on generating a total of 625 models and averaging to make a final model for each β-tubulin isotype.

4.3.3. *In silico* docking of TBZ into β-tubulins

Potential binding sites were evaluated using the Site Finder application and recent computational work on benzimidazole binding to parasitic β-tubulins (Aguayo-Ortiz *et al.* 2013). Conformational variants of TBZ were created in 3-D within MOE. A database of conformations was then used to dock TBZ to the wild-type homology model using induced fit and template similarity protocols. The placement was scored with Triangle Matcher and rescored with London dG. Poses were refined with the Amber10:EHT forcefield with GVBI/WSA dG scoring. Candidate poses were then identified by inspecting polar contacts. Geometry optimization was carried out with MOPAC 7.0 using AM1. Conformational analysis of the bound structure was evaluated with LowModeMD (Labute 2010). 2-D contact maps were created using Ligand Interactions (Clark and Labute 2007).

4.3.4. Cell culture

HUVEC cells were purchased from Clonetics and were used between passages 4 and 9. HUVECs were cultured on 0.1% gelatin-coated (Sigma) plates in endothelial growth medium-2 (EGM-2; Clonetics) in tissue culture flasks at 37 °C in a humidified atmosphere of 5% CO_2. NIH-3T3 cells were obtained from Vishy Iyer at the University of Texas at Austin and cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 10% bovine calf serum.

4.3.5. Immunohistochemistry

Cell lines were cultured in 6-well plates and treated with thiabendazole dissolved in 1% DMSO. Control cells received 1% DMSO. After 24 h, cells were fixed with methanol at -20 °C for 10 min and subsequently with 4% paraformaldehyde in PBS at room temperature for 10 min. Cell membranes were permeabilized with 0.2% Triton X-100 in PBS, and nonspecific antibody binding sites were blocked with 5% goat serum for 1 h at room temperature. Cells were incubated with primary antibodies to EB1 (BD Bioscience) and α-tubulin (Sigma) at 4 °C overnight. After washing with PBST, primary antibodies were detected by Alexa Fluor-488 or 555 goat anti-rabbit or mouse immunoglobulin (IgG). 4′,6-Diamidino-2-phenylindole (DAPI dye, Sigma) was added as needed to visualize nuclei.

4.3.6. Cell transfection and perfusion

EB3-eGFP cDNA obtained from Anna Akhmanova was cloned into the vector CS2+(Stepanova *et al.* 2003). TUBB4 (Origene, RG203945) and TUBB8 (Origene, RG213889) cDNAs were purchased and cloned into the vector CS107-RFP-3Stop.

HUVEC cells were transfected by nucleofection (Lonza) according to the manufacturer's instructions. To analyze the effect of TBZ in living cells, we used a closed perfusion system (POC-R2, Pecon) connected to a peristaltic pump (Ismatec). 1% DMSO, 250 μM TBZ or 1% DMSO diluted in EBM-2 medium was flowed at 100 μl/min rate for the indicated times.

4.3.7. Western blotting

HUVECs were cultured in 6-well plates and treated with 1% DMSO or 1% DMSO, 250 μM TBZ for 24 hours. Cells were lysed in cell lysis buffer (Cell Signaling Technology) containing 1 mM PMSF and analyzed by SDS-PAGE and western blotting using anti-EB1 (BD Bioscience) or anti-EB3 (Millipore) or anti-Clip170 (Santa Cruz) antibodies.

4.3.8. Imaging and image analysis

Immunohistochemistry experiments, live HUVECs, and live *KDR:GFP* transgenic *Xenopus laevis* were imaged using an inverted Zeiss LSM5 Pascal and Zeiss LSM700 confocal microscope, and super-resolution structured illumination (SR-SIM) combined with Zeiss LSM710 microscope. Comet lengths were measured using the software Fiji. Confocal images were cropped and enhanced in Adobe Illustrator and Adobe Photoshop for the compilation of figures.

4.3.9. Benzimidazole clustering analysis

81 commercially used benzimidazole compounds spanning a wide range of classes were curated from PubChem(Kim *et al.* 2016). JOElib

(http://joelib.sourceforge.net), OpenBabel (O'Boyle *et al.* 2011), and Chem Mine features were computed using ChemMine tools (Backman *et al.* 2011). Heatmaps were visualized using Morpheus (https://software.broadinstitute.org/morpheus). Clustergrams were generated by hierarchical clustering on the one minus Pearson correlation coefficient with average linkage.

4.3.10. Humanizing yeast β-tubulin using CRISPR-Cas9

The human TUBB4 and TUBB8 open reading frames were integrated chromosomally (from start to stop codon) into *Saccharomyces cerevisiae* in place of the endogenous *TUB2* open reading frame using CRISPR/Cas9 genome editing as described previously. (Akhmetov *et al.* 2018a). Two sgRNAs were designed targeting the yeast *TUB2* locus using the Geneious (v10.2.6) CRISPR-Cas9 tools suite, purchased as oligos from IDT, and cloned into yeast CRISPR-K/O vectors using the yeast toolkit (YTK) (Lee *et al.* 2015) to express a synthetic guide RNA sequence, Cas9 nuclease, and a selectable marker (URA3) (Akhmetov *et al.* 2018b; Garge *et al.* 2020). Repair templates were constructed by PCR amplification of the human β-tubulin ORF (from the human ORFeome (Lamesch *et al.* 2007)) flanked by 75 bp of target chromosomal boundary at the *TUB2* locus to facilitate recombination *via* homology directed repair. BY4741 (S288C) yeast strains were co-transformed with the CRISPR/Cas9 vector and repair template using Zymo Research Frozen-EZ Yeast Transformation II Kit. Transformants were selected on SC-URA media. Surviving colonies were screened by colony PCR, and Sanger sequenced to confirm replacement.

4.3.11. Humanized yeast growth assays

Assayed benzimidazole compounds were all dissolved in 100% DMSO to prepare stock solutions of 5 or 10 mg/ml based on solubility. Candidate VDA compounds were titrated in ranges of 5-1000 µg/ml into growth medium depending on solubility (**Fig. 4.6.S5** lists specific concentrations) for subsequent growth assays. Liquid growth assays were performed in triplicate in 96-well format using a Biotek Synergy HT incubating spectrophotometer. Humanized tubulin strains were pre-cultured to saturation in YPD and diluted into 150 µL of media to have 0.05-0.1 x 10^7 cells/ml. Assays were typically run for 48 hrs with absorbance measured every 15 min.

4.3.12. *Xenopus* embryo manipulations and VDA assays

Xenopus embryos were reared in 1/3× Marc's modified Ringer's (MMR) solution. Each drug was treated to embryos from stage 31 until stage 38 with 10 µg/ml or 20 µg/ml in 1% DMSO diluted in 1/3X MMR. Embryos were fixed at stage 38 with MEMFA, and whole mount *in situ* hybridization for *erg* was performed as described in Sive *et al.*[87].

4.4. RESULTS

4.4.1. Thiabendazole disrupts microtubule plus ends in endothelial cells

Thiabendazole exhibits broad-spectrum activity against fungal and nematode crop pests (Lubega and Prichard 1990), but prior to demonstration of its VDA activity, it was generally thought to lack activity in tetrapods (EPA 2002). Its binding target and

mechanism of vascular action remains poorly understood (Lacey and Gill 1994), although *in vitro* studies have suggested various benzimidazole compounds inhibit cell growth by interfering with microtubule polymerization (Davidse and Flach 1978; Lacey and Gill 1994; Aguayo-Ortiz *et al.* 2013; Vela-Corcía *et al.* 2018). Benzimidazole suppressor screens in both *Saccharomyces cerevisiae* and *Caenorhabditis elegans* have independently identified resistance mutations occuring in β-tubulin genes, giving some insight into the binding site (Driscoll *et al.* 1989; Hahnel *et al.* 2018). The case for a β-tubulin binding site is strengthened by numerous animal and agricultural studies also demonstrating resistance mutations arising repeatedly and independently across multiple parasitic nematode and fungal species infecting farm livestock and crops (Brown *et al.* 1984; Tolliver *et al.* 1993; Nakaune and Nakano 2007; Zhang *et al.* 2009; Skuce *et al.* 2010; Carter *et al.* 2013; Aguayo-Ortiz *et al.* 2013; Furtado *et al.* 2014; Furtado and Rabelo 2015; Zhu *et al.* 2016; Ramünke *et al.* 2016; Yilmaz *et al.* 2017; Yang *et al.* 2018; Ali *et al.* 2018; Baltrušis *et al.* 2018).

To examine the effects of the microtubule cytoskeleton and dynamics in the presence of TBZ, we examined the localization of GFP-tagged EB1, which labels growing microtubule plus ends and provides a proxy for microtubule dynamics in both endothelial (HUVEC) and non-endothelial (NIH3T3) human cells. Despite the grossly normal architecture of microtubules in TBZ-treated HUVECs, TBZ significantly reduced the accumulation of EB1 at microtubule plus ends (**Fig. 4.6.1A', B'**) as compared to its control (**Fig. 4.6.1A, B**). Importantly, and consistent with the overall normal morphology and patterning of TBZ-treated embryos (Cha *et al.* 2012), we found that TBZ had a substantially less robust effect on EB1 accumulation at microtubule plus ends in

fibroblasts as compared to endothelial cells (**Fig. 4.6.1C-E**). These data are consistent with, and provide new insights into, previous studies showing that TBZ's interaction with tubulin interferes with microtubule polymerization in nematodes and fungi (Dawson *et al.* 1984; Davidse 1986).

4.4.2. Thiabendazole selectively targets TUBB8 among human β-tubulins

Three commonly observed mutations in fungal and nematode β-tubulins (F200Y, E198A, and F167Y) confer resistance to TBZ (**Fig. 4.6.2A, B**), suggesting that its binding site is in the vicinity of these residues (Brown *et al.* 1984; Tolliver *et al.* 1993; Nakaune and Nakano 2007; Zhang *et al.* 2009; Skuce *et al.* 2010; Carter *et al.* 2013; Aguayo-Ortiz *et al.* 2013; Furtado *et al.* 2014; Furtado and Rabelo 2015; Zhu *et al.* 2016; Ramünke *et al.* 2016; Yilmaz *et al.* 2017; Yang *et al.* 2018; Ali *et al.* 2018; Baltrušis *et al.* 2018) (**Table 4.7.1**). Based on the previously observed benzimidazole suppressor mutations, we used 3D structural modeling to evaluate TBZ's potential binding sites in a fungal β-tubulin. We first constructed 3D homology models of the *Schizosaccharomyces pombe* (fission yeast) wild-type and TBZ-resistant F200Y β-tubulins, based on the previously determined *Ovis aries* β-tubulin crystal structures (PDB: 3UT5 (Ranaivoson *et al.* 2012) and 3N2G (Barbier *et al.* 2010)) as templates. We computationally refined the structures and then evaluated potential binding modes of TBZ, as detailed in the Methods, using computational docking algorithms to localize TBZ's potential binding sites within the fungal β-tubulin structures (**Fig. 4.6.S1**). We identified a binding site around F200 to be the most probable (**Table 4.7.2**). We found that the preferred binding conformations of TBZ in both models (**Fig. 4.6.S1**) situated close to (but distinct from) the colchicine binding site. These observations were in strong agreement with

computational predictions made on parasitic β-tubulins binding benzimidazoles (Aguayo-Ortiz *et al.* 2013) and recent crystal structures of other benzimidazole derivatives binding to bovine brain β-tubulins (Wang *et al.* 2016) .

On measuring the polar contacts and clashing energies of TBZ with tubulin, we found that the wild-type β-tubulin bound to TBZ more favorably with contact energy (-9.9 kcal/mol) as compared to its F200Y counterpart, which showed unfavorable repulsions (+27.6 kcal/mol) (**Table 4.7.2, S3**). For the wild-type protein, TBZ's polar contacts included E198 and Q134 (**Fig. 4.6.2C, S1**). Arene-hydrogen interactions between the drug and protein included contributions from F200, L250, and L253. However, for our F200Y mutant, repulsion was observed in our fixed ligand experiments predominantly caused by unfavorable contacts made with Y200, F240, L250, and L253 (**Fig. 4.6.S1**). Our analyses suggest F200Y likely forms a hydrogen bond to E198 in the TBZ-resistant mutant, thus constricting the pocket and occluding binding.

Unlike fungi, tetrapods have multiple β-tubulin isotypes (here, we use the term isotype to denote the protein products of paralogous genes, in accordance with prior tubulin literature), and their expression varies in different cells and tissue types. For example, human tubulin βI (TUBB/TUBB5) is constitutively expressed in many cells and tissues, whereas βIII (TUBB3) is exclusively enriched in neurons and the brain (Kavallaris 2010; Leandro-García *et al.* 2010). The specific roles of different β-tubulin isotypes are not yet fully understood, but recent studies indicate that their sequence diversity modulates binding affinity to tubulin-binding drugs and influences microtubule dynamics through distinct interactions with molecular motors (Kavallaris 2010;

Sirajuddin *et al.* 2014). The recurrence of TBZ resistance mutations at the same three loci across diverse fungi and nematodes (**Table 4.7.1**) led us to hypothesize that human β-tubulin isotypes might have differential sensitivities to TBZ by virtue of incorporating resistant residues at positions 167, 198, and 200, potentially explaining both its tissue-specific effects and generally low toxicity in humans.

Indeed, multiple sequence alignment of human and yeast β-tubulin genes indicated that while F167 remained conserved across all the human isotypes, positions 198 and 200 were variable (**Fig. 4.6.2B**). Moreover, all human β-tubulin isotypes except TUBB1 and TUBB8 contain the F200Y resistance mutation. Because TUBB1 also harbors the other commonly observed E198A suppressor (**Fig. 4.6.2B**), TUBB8 is the only human β-tubulin isotype predicted by sequence to be TBZ-sensitive. Given this variability across isotypes, we next asked how the E198A and F200Y mutations would be expected to affect TBZ's ability to bind at its predicted site in human isotypes.

We first evaluated this hypothesis computationally, by constructing 3D homology models for each of the human β-tubulin isotypes in the same manner as for the fungal model (see Methods). We then performed induced-fit docking with TBZ across our human β-tubulin models. Using a TBZ-wild-type fungal β-tubulin complex as a template, we docked TBZ into the same pocket in each of the human isotypes and measured protein-ligand interactions in the superimposed structures. In agreement with our primary sequence based predictions, TBZ fit well into the predicted binding pocket of only TUBB1 and TUBB8, which lack the F200Y mutation. Both showed favorable binding energies of -1.8 and -8.3 kcal/mol, respectively (**Table 4.7.2**). The large difference in

contact energy among these isotypes could be explained by position 198. In TUBB1, alanine occupies position 198, whereas TUBB8 has glutamate, which contributed heavily to the binding energy in all of our simulations when both F200 and E198 were present. Our data suggest that TBZ binding is stabilized by hydrogen bonds with residues Q134 and E198 in the presence of F200 (**Fig. 4.6.S1, S2**). Taken together, our *in silico* studies predicted that TBZ should strongly bind TUBB8 and weakly bind TUBB1, but should not bind any other human β-tubulin isotypes.

4.4.3. Functional assays in human endothelial cells and humanized yeast confirm TBZ specificity to human TUBB8

Given TBZ's effects on human vascular endothelial cells and *in vivo* vascular disruption in *Xenopus* embryos (Cha *et al.* 2012), we wished to test directly if TUBB8-specific binding could explain the compound's effects. We thus asked whether resistance to TBZ could be acquired by simply supplying human β-tubulin isotypes predicted to be resistant. We tested this by two independent assays: (i) by overexpressing specific sensitive or resistant human β-tubulin isotypes in human endothelial cells and (ii) by humanizing Baker's yeast's β-tubulin *TUB2* to enable assays of individual human β-tubulin isotypes.

To test if microtubule dynamics in human cells could be significantly restored by supplying resistant β-tubulin isotypes, we singly transfected HUVEC cells with plasmids overexpressing either TUBB4 or TUBB8 and assayed microtubule dynamics by measuring the comet lengths of end-binding protein EB3 (**Fig. 4.6.3A**). Compared to untransfected HUVECs, we saw that overexpressing TUBB4 significantly rescued the decrease in comet length observed in TBZ-treated cells (**Fig. 4.6.3B**). Transfection with

TUBB8, by contrast, had no effect (**Fig. 4.6.3B**). The differences became very significant after 30 minutes of exposure (**Fig. 4.6.3B**).

As an independent assay of TBZ action on human tubulins, we turned to humanized yeast, as our previous work showed that of the nine human β-tubulins, only TUBB4 and TUBB8 could functionally replace *TUB2* in *Saccharomyces cerevisiae* (Garge *et al.* 2020). From our modeling and docking data, we hypothesized that yeast strains humanized with TUBB8 would be susceptible to TBZ while humanizing with TUBB4 would confer TBZ resistance. *Saccharomyces cerevisiae* possesses 2 α-tubulins (*TUB1* and *TUB3*) that interact with *TUB2* to form tubulin heterodimers, which in turn oligomerize to form microtubules. Wild-type BY4741 haploid strains are TBZ-resistant. However, previous studies have shown that on deleting *TUB3*, yeast strains become susceptible to benzimidazoles (Schatz *et al.* 1986b) likely due to reduced overall α-tubulin stoichiometry or possibly by TBZ occluding *TUB2*'s dimerization with *TUB1* but not *TUB3*. Therefore, we performed all our yeast replacement assays in a *tub3Δ* background, which yielded a clear growth defect in the presence of TBZ (**Fig. 4.6.S3**). In order to test the effect of TBZ on human β-tubulin isotypes TUBB4 and TUBB8, we used CRISPR/Cas9 to construct yeast strains with these human isotypes in place of the endogenous *TUB2* and tested them in the presence or the absence of the drug (**Fig. 4.6.3A**). We found that strains possessing wild-type *TUB2* and human TUBB8 exhibited slow growth in the presence of TBZ (at conc. 20 μg/ml). By contrast, the strain humanized with TUBB4, which is predicted to be resistant to TBZ, grew normally in the presence of TBZ (**Fig. 4.6.3C**, **S3A**).

Together with our *in silico* docking data, our results in HUVECs and humanized yeast indicate that TUBB8 is uniquely TBZ-sensitive, suggesting in turn that vascular endothelial cells are selectively sensitive to its loss.

4.4.4. Benzimidazole resistance patterns and chemical similarities suggest additional VDAs

Given the plethora of fungal and nematode studies on benzimidazole pesticide resistance in agriculture (Jambre *et al.* 1979; Middelberg and McKenna 1983; Brown *et al.* 1984; Dawson *et al.* 1984; Tolliver *et al.* 1993; Romero and Sutton 1998; Gossen *et al.* 2001; Saeed *et al.* 2007; Nakaune and Nakano 2007; Banno *et al.* 2008; Cabañas *et al.* 2009; Zhang *et al.* 2009, 2016a; b; Skuce *et al.* 2010; Niciura *et al.* 2012; Xu *et al.* 2015; Zhu *et al.* 2016; Kumar *et al.* 2016; Liu *et al.* 2016, 2018; Chagas *et al.* 2016; Ramünke *et al.* 2016; Rupp *et al.* 2017; Keegan *et al.* 2017; Yilmaz *et al.* 2017; Santos *et al.* 2017; Yang *et al.* 2018; Ali *et al.* 2018) (**Fig. 4.6.4A**), we reasoned that TBZ's molecular mechanism may extend to other commercially used benzimidazole compounds. Indeed, based on our experiments, a simple epidemiological signature should be sufficient to identify other pesticides that likely to function as vascular disrupting agents and angiogenesis inhibitors: (i) the compounds should be selectively toxic to fungal and nematode clades but demonstrate low toxicity in tetrapods, and (ii) sensitive species should specifically gain benzimidazole resistance from F167Y, E198A or F200Y β-tubulin mutations.

In order to understand how extensively distributed benzimidazole resistance was, we mined ~40 years of literature to identify reported cases of pesticide resistant species seen in wild and parasitic nematodes and fungi. Benzimidazole resistance is a global

phenomenon (**Fig. 4.6.4A**); across 9 major commercial benzimidazole-based pesticides, we found multiple independent instances of reported resistance across 27 (12 nematodes and 15 fungal) parasitic species (**Table 4.7.1**), all of which exhibited at least 1 of the 3 signature β-tubulin mutations. These widespread patterns of benzimidazole pesticide resistance suggested at least 9 new candidate VDAs.

As a complement to the epidemiological data, we also considered chemical properties by asking if pesticide benzimidazoles shared similar chemical feature profiles relative to other benzimidazoles. We curated >80 commercially available compounds in the benzimidazole class spanning a diverse range including pesticides, fungicides, therapeutics, and preservatives. Upon hierarchical clustering of these benzimidazoles based on their chemical properties computed from JOELib's features matrix (Cao *et al.* 2008; Backman *et al.* 2011) (**Table 4.7.5**), we found that pesticide benzimidazoles generally shared similar chemical properties and clustered together (**Fig. 4.6.4B**).

4.4.5. Numerous commercially used benzimidazoles also function as vascular disrupting agents

We next tested if pesticides exhibiting the epidemiological signature and clustering in the same clades by virtue of their chemical features would also specifically inhibit TUBB8 and function as VDAs. We selected 12 commercially used benzimidazole compounds across 2 clusters (**Fig. 4.6.4B**). Our list included 2 anthelmintics, both World Health Organization essential medicines (albendazole and mebendazole) prescribed to treat broad-spectrum human intestinal nematode infections; fenbendazole, an anthelmintic prescribed specifically for animals against gastrointestinal nematode parasites; 2 currently banned pesticides, benomyl and carbendazim, formerly used in

agriculture; triclabendazole, specifically used to treat liver fluke infections; and 5 proton-pump inhibitors (esomeprazole, lansoprazole, omeprazole, pantoprazole, and rabeprazole) used to treat gastrointestinal and stomach acid disorders. The latter set were from a different clade and did not exhibit the epidemiological signature, serving as negative controls.

We first took advantage of our humanized yeast strains to rapidly discriminate TUBB8-specific inhibition from general β-tubulin inhibition. We found that 5 of the 12 compounds tested selectively inhibited TUBB8, as evidenced by the growth profiles observed for the humanized strains when cultured in the presence of the drugs (**Fig. 4.6.5, S4**). Notably, none of the 5 proton pump inhibitors or colchicine exhibited any tubulin inhibition (**Fig. 4.6.S4, S5A**), confirming the specificity of the epidemiological signature as a predictor of TUBB8 inhibition. In contrast, triclabendazole was generally toxic, behaving as a pan-isotype inhibitor (**Fig. 4.6.S5B**).

Testing the 5 positive TUBB8-inhibiting compounds in *Xenopus laevis* embryos showed strong vascular disrupting activity for all 5 compounds (**Fig. 4.6.5**). As we observed previously for TBZ (Cha *et al.* 2012), the gross morphology of the treated embryos was largely normal (**Fig. 4.6.5)**. Thus, this broader class of benzimidazoles do in fact generally act as vascular disrupting agents in vertebrates.

4.5. DISCUSSION

In the >30 years of therapeutic research efforts in the angiogenesis field only a highly restricted set of drugs have yet been approved (Hinnen and Eskens 2007). Given the frequent failure to successfully make it through clinical trials and the high costs and lengthy process associated with developing new compounds, drug repurposing can offer efficient alternatives in developing new patient therapies with accelerated timeframes. This study represents a rather unconventional path to drug repurposing, leveraging a combination of model organisms, humanized yeast, cell culture, molecular modeling, and epidemiological data mining to determine TBZ's molecular target and mechanism of vascular action. Indeed, TBZ was initially identified as a VDA and angiogenesis inhibitor by using a Baker's yeast model of angiogenesis discovered in a computational search for orthologous phenotypes, or phenologs, aimed at exploiting deep evolutionary conservation to prioritize yeast processes relevant to human diseases (McGary *et al.* 2010; Cha *et al.* 2012; Woods *et al.* 2013). Although obviously lacking blood vessels and a circulatory system, yeast nonetheless retains conserved biological pathways and processes relevant to vertebrate angiogenesis genes, and it was on the basis of these conserved processes that the antifungal compound TBZ was initially suspected, later confirmed, to be an angiogenesis inhibitor (Cha *et al.* 2012).

While TBZ somewhat reduced the abundance of tubulin proteins in human cells (Cha *et al.* 2012), at angiogenesis-inhibiting doses, the overall morphology of TBZ treated animals was normal, suggesting that only certain cell types, specifically those endothelial cells involved in forming the vasculature, might be uniquely susceptible to

TBZ. Here, we find that TBZ does indeed specifically modulate the microtubules in vascular endothelial cells. Several currently identified microtubule targeting drugs have been reported to interfere with polymerization dynamics by binding β-tubulin (Tozer *et al.* 2005; Hinnen and Eskens 2007) close to or at the colchicine binding site. Building on previous work(Aguayo-Ortiz *et al.* 2013; Wang *et al.* 2016), our *in silico* modeling results suggest that TBZ's binding site, while in close proximity to the colchicine binding site, is distinct from it, thereby uncovering a novel β-tubulin effector site likely specific to other benzimidazoles and TBZ analogs.

In contrast to β-tubulin anticancer drugs, which have largely shown pan-isotype activity, to our knowledge, this study presents an unusual case of isotype-specific drug targeting in the β-tubulin gene family. Fungal suppressor studies on benzimidazole resistance have repeatedly found resistant mutations in β-tubulin; we found that 8 of 9 human β-tubulins natively harbor the same suppressor mutations and consequently exhibit unfavourable steric clashes interfering with TBZ binding. We demonstrate both *via* human cell culture microtubule assays and humanized yeast drug sensitivity tests that TBZ selectively targets only TUBB8 among the nine human β-tubulins, thus disrupting microtubule dynamics and reducing end-binding protein accumulation at the plus ends of microtubules in susceptible cells.

With TUBB8 thus acting as the specific target, it follows that of all human cell types, vascular endothelial cells must in turn be particularly sensitive to inhibition of TUBB8, leading to selective disruption of the vasculature relative to other human tissues. It remains to be seen why TBZ's vascular disrupting activity is restricted to immature or

newly forming blood vessels, but we speculate that this subset of the vasculature lacks reinforcing cell-cell contacts typical of larger, more established vasculature, leading to greater sensitivity to TBZ-induced microtubule disruption. As β-tubulin isotypes tend to be broadly expressed and often substitute for one another in microtubule structures (Uhlén *et al.* 2015; The UniProt Consortium 2019), one possibility is that TUBB8 inhibition simply leads to the loss of interactions with endothelial cell-specific components, thus specifically impacting vasculogenesis/angiogenesis. However, gene-gene and gene-drug interactions can often proceed by less obviously direct mechanisms to selectively impact cell types or phenotype penetrance *via* conditional cell-specific or dosage-dependent synthetic interactions (Burga *et al.* 2011; O'Neil *et al.* 2017). It would thus not be surprising for the consequences of inhibiting TUBB8 in vascular endothelial cells to be similarly indirectly mediated by endothelial cell-specific synthetic interactions. Further experiments characterizing TBZ's selective activity against newly forming/formed vasculature and the vascular-specific roles of TUBB8 in tetrapods could offer valuable insights into the cytoskeletal dynamics underlying vasculogenesis and angiogenesis.

Based on chemical properties and signature resistance mutations observed against benzimidazole compounds, we identified a larger class of extensively used fungicides and pesticides that all exhibit vascular disruption activity. While our results suggest possible new clinical applications for these compounds, they also highlight the potential caveats of their use in at-risk populations, especially for the two compounds (albendazole and mebendazole) that are FDA approved for human use. The WHO recommends the use of both albendazole and mebendazole as essential antihelmenthics worldwide for children

up to the age of 14 against soil-transmitted helminth infections. Moreover, these compounds are widely used as public health interventions in pregnant women after the first trimester in regions where hookworm and whipworm infections exceed 20% (World Health Organization 2018). While in the US, the risk of mebendazole use during pregnancy has not been assigned, our data add weight to WHO recommendations that these drugs should not be administered in the first trimester of pregnancy and suggest their use be carefully evaluated in patients in which angiogenesis inhibition might pose risks, including using caution later in pregnancy in light of the evidence that the compounds disrupt immature vasculature and might prove harmful to a developing fetus. Conversely, while efforts in the angiogenesis field have been often motivated towards developing anticancer therapies, the wide use of the compounds discussed here and their FDA-approved status could open alternative paths to treating other angiogenesis and/or vascular related diseases, such as diabetic retinopathy, macular degeneration, and hemangioma. It remains to be seen if other benzimidazoles sharing similar chemical profiles to those tested in our work (such as ciclobendazole, nocodazole, oxibendazole, and oxfendazole) also exhibit vascular disrupting activity.

More broadly, our framework of leveraging phenotypic relationships between species and repurposing model organisms to systematically explore drug mechanisms opens new routes for drug repurposing and discovery and highlights the power of systems biology and evolution-guided approaches in advancing our knowledge of conserved genetic modules and how their disruption manifests in disease. This work also illustrates how duplicated genes diversify their functions and reinforces the therapeutic benefits of finding drugs specific to individual gene family members. As evidenced by the high

degree of replaceability of conserved genes from cross-species complementation assays (Kachroo *et al.* 2015; Laurent *et al.* 2016, 2020), we anticipate that the combination of humanized yeast and phenolog-based disease modeling can be extended beyond vascular disruption to other conserved processes and therapies targeting them.

4.6. FIGURES

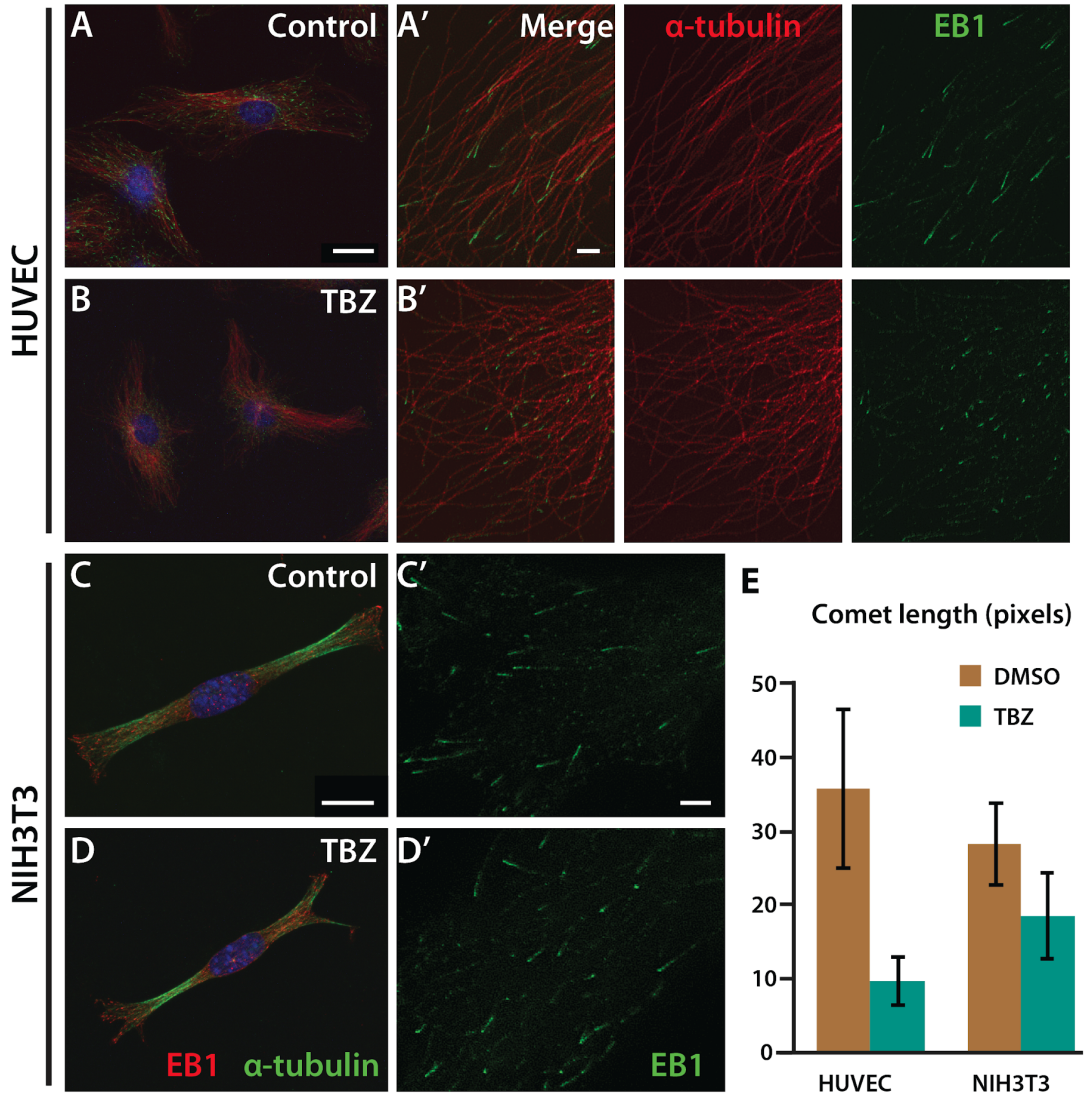


Figure 4.6.1. Thiabendazole (TBZ) significantly reduces EB1 comet length at microtubule plus ends in cultured human cells.

Immunohistochemical analysis of α-tubulin in two human cell lines using confocal microscopy does not show a definite distinction between 1% DMSO-treated control (**A,**

C) and 1% DMSO, 250 μM TBZ-treated cell lines (**B, D**), but images from super-resolution microscopy reveal that the accumulation of end-binding (EB) protein 1 at the plus end of microtubules is significantly reduced with TBZ treatment (**B**) compared to the control (**A**) in HUVECs. In NIH-3T3 cells, the reduced EB1 comet length following TBZ treatment (**D**) compared to control (**C**) is not as pronounced as in HUVECs, as quantified by comet length (**E**). Scale bars, 20 μm in (A) and (C), 2 μm in (A') and (C').

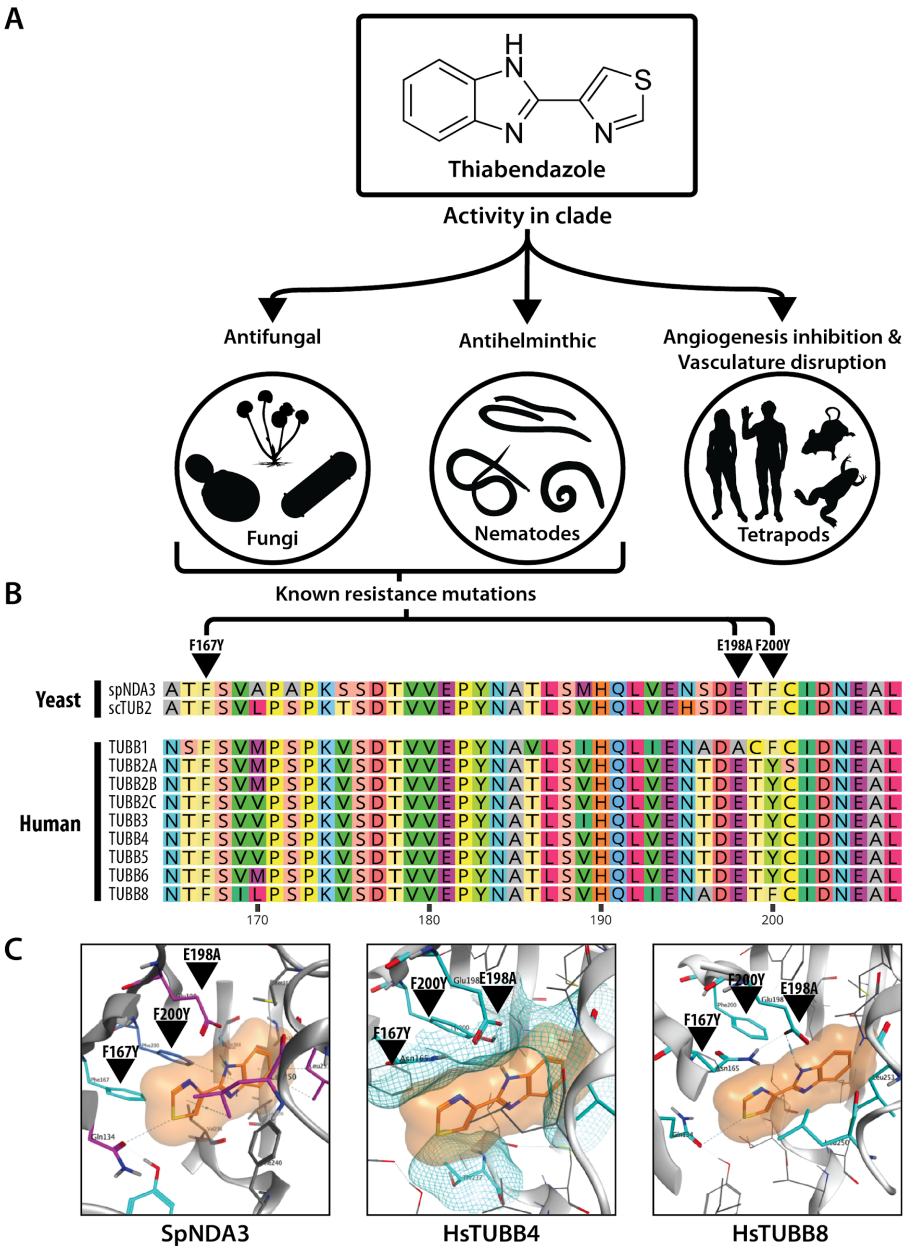


Figure 4.6.2. Uncovering the molecular mechanism of thiabendazole.

(**A**) TBZ elicits varying activity across different clades of life being toxic to fungal and nematode clades but behaves as a vascular disrupting agent in tetrapods. (**B**) Of the 9

human β-tubulins, 8 have amino acids at positions 167, 198, and 200 that confer TBZ resistance to fungal tubulins , as seen in a multiple sequence alignment of human and *Schizosaccharomyces pombe* β-tubulins; only TUBB8 lacks resistance mutations. (**C**) *In silico* docking of TBZ (orange) into a homology modeled yeast β-tubulin 3D structure (see Methods) indicates TBZ is well-accommodated by a binding pocket in wild-type yeast *NDA3* that abuts the 3 major β-tubulin TBZ resistance mutation sites. In contrast, docking of TBZ into homology models of human TUBB4 and TUBB8 indicates the potential for differential binding, with TUBB8 accommodating TBZ whereas, in the case of TUBB4, TBZ is sterically blocked. Polar contacts are illustrated *via* dashed lines, and residues lining the proposed binding pocket are shown in cyan. Intramolecular hydrogen bonding between E198 and Y200 in TUBB4 reorganizes the geometry of the binding pocket. Residues involved in steric clashing are depicted with a partial mesh surface. (Note that due to steric clashes between TBZ and TUBB4 at the proposed binding pocket, TBZ was superimposed from our binding model to measure interactions).

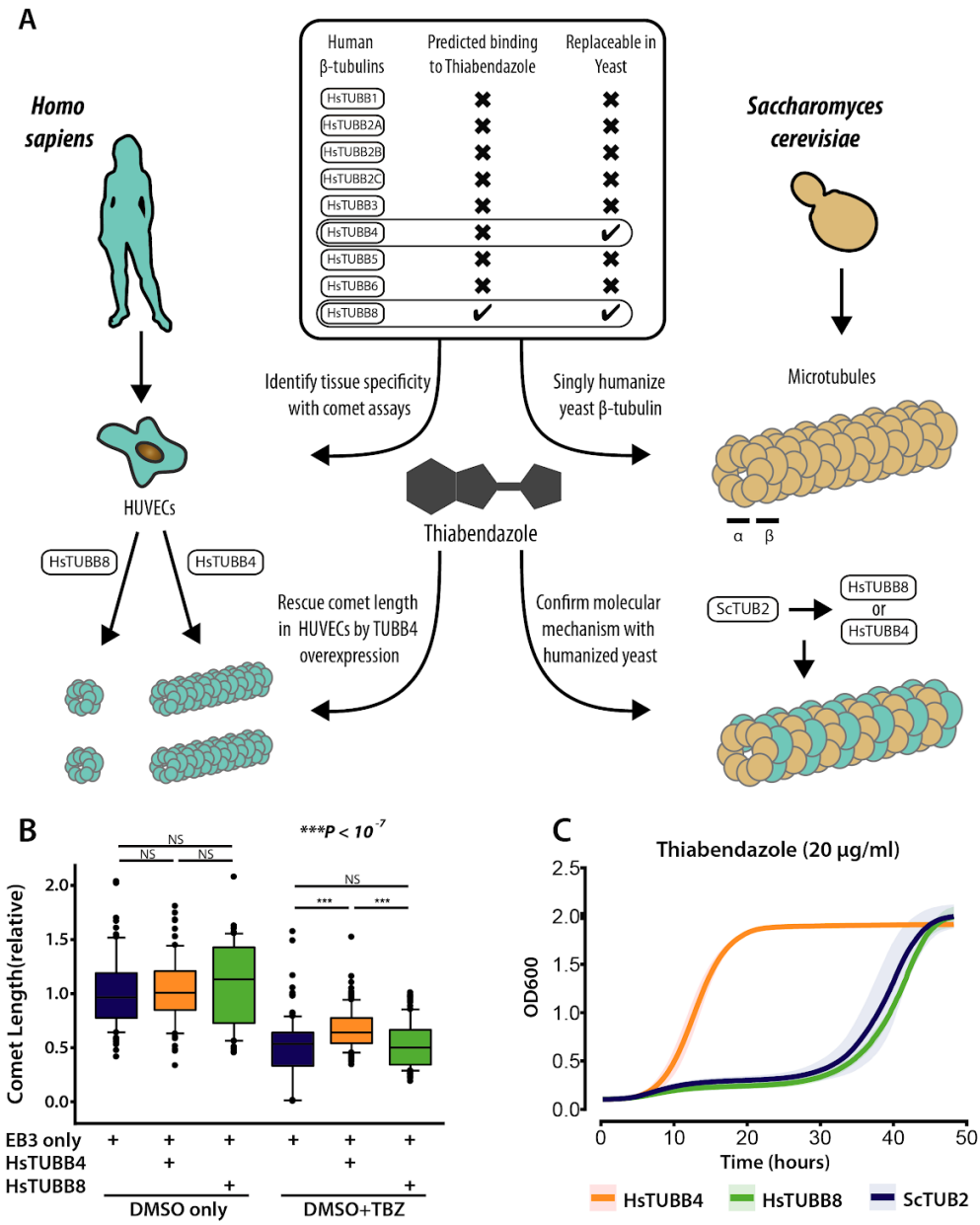


Figure 4.6.3. TBZ specifically inhibits the human β-tubulin TUBB8, not TUBB4, in humanized yeast and HUVEC cell culture.

(**A**) Overview. TBZ's isotype specificity was identified in 2 ways. (Left) Recombinant human β-tubulins TUBB4 and TUBB8 were individually overexpressed in HUVEC cell

culture to monitor comet lengths in the presence of TBZ. (Right) Using humanized yeast wherein yeast *TUB2* was singly humanized by either of 2 replaceable human β-tubulins TUBB4 or TUBB8 to screen for differential sensitivity towards TBZ. (**B**) Reduced EB3 comet length after 1% DMSO, 250 μM TBZ treatment compared to 1% DMSO treated control. **A.** Comet length is similar in EB3, TUBB8 transfected HUVECs compared to EB3 transfected controls expressing native tubulins, but comets are longer in most EB3, TUBB4 transfected cells. **B.** Comet length is statistically similar between cells treated with 1% DMSO; however, following 30 minutes of 1% DMSO, 250 μM TBZ treatment *TUBB4* transfected cells have significantly longer EB3 comets than HUVECs with TUBB8 or expressing native tubulins. (**C**) Growth profiles of humanized yeast strains show TBZ's isotype specificity to TUBB8. When grown in the presence of TBZ, Strains carrying the wild-type *TUB2* (blue) and human TUBB8 (green) genes are sensitive to TBZ while humanized TUBB4 strains (orange) are resistant. Mean +/- standard deviation indicated by solid lines and shaded boundaries, respectively.

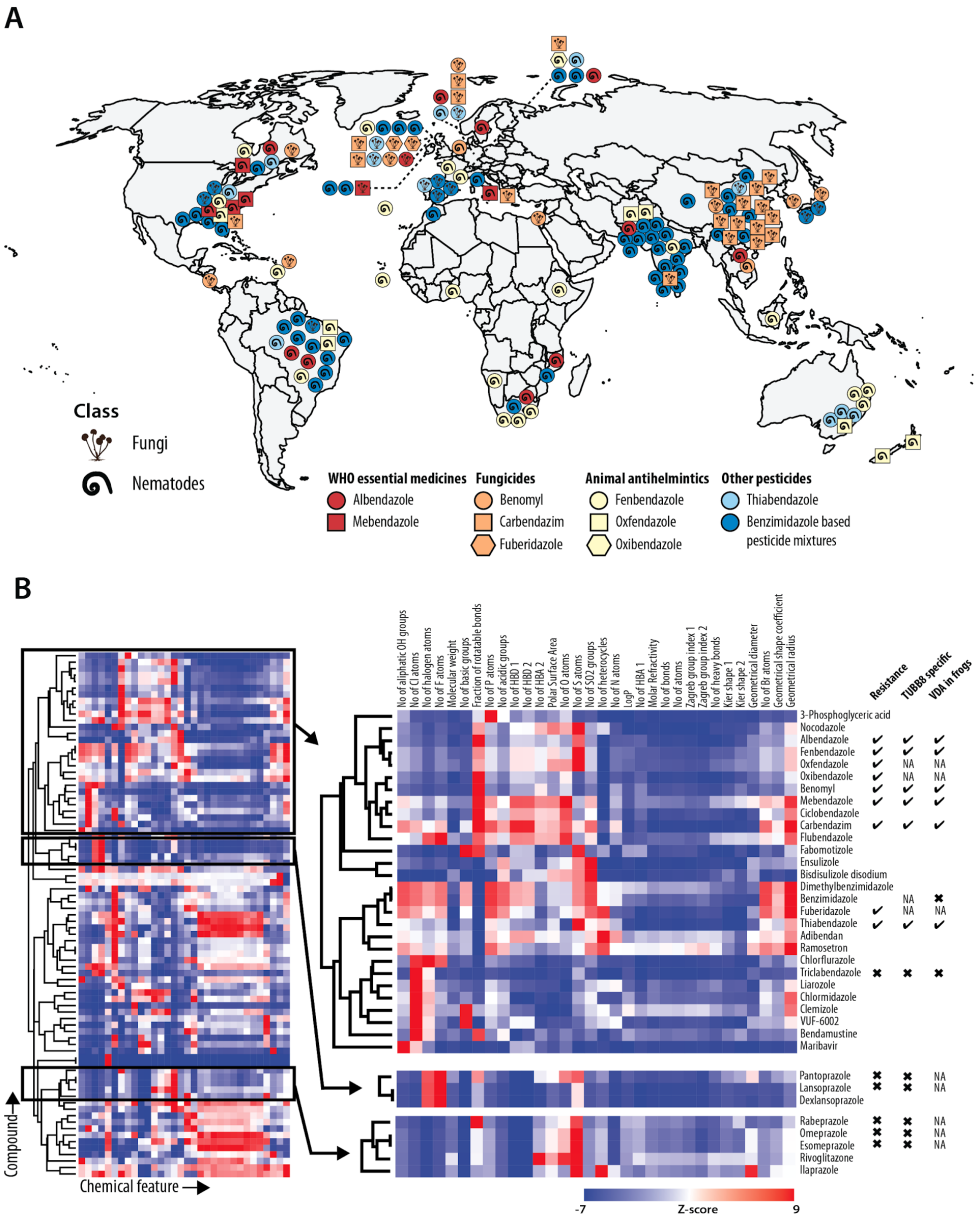


Figure 4.6.4. Global trends in benzimidazole resistance mutations and chemical structural similarities suggest numerous potential vascular disrupting agents.

(**A**) 3 β-tubulin mutations, (F167Y, E198A, F200Y) conferring benzimidazole resistance have been globally observed among parasitic nematode and fungal species. Each icon

represents an instance of β-tubulin suppressor mutations occurring in benzimidazole resistant parasitic fungal or nematode species (See **Table 4.7.1** for list of species showing benzimidazole resistance). (**B**) Commonly used benzimidazoles hierarchically clustered by their chemical properties suggest new vascular disrupting agents with similar molecular mechanisms to TBZ. (Left) Clustergram of 81 widely used benzimidazole compounds spanning a wide range of drug classes grouped by chemical features (See **Table 4.7.5** for the full list of compounds and features analysed). (Right) Zooms of black boxes indicate 3 clades containing TUBB8 specific VDA candidates (top) and proton-pump inhibitors (bottom).

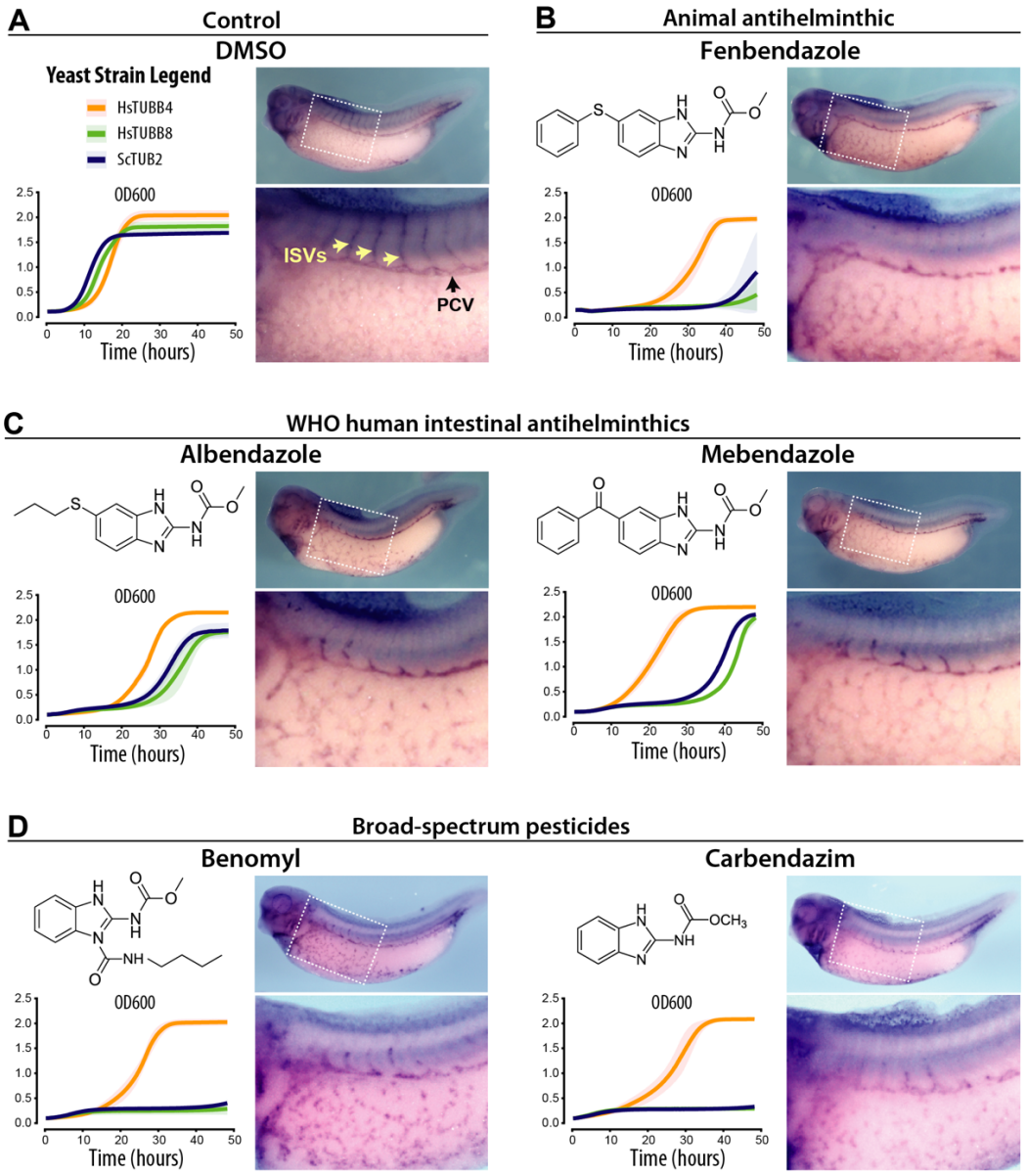


Figure 4.6.5. Commercially used benzimidazole pesticides, antifungals, and antihelminthics are also TUBB8-specific and disrupt vasculature.

(**A-D**) *In situ* hybridization of blood vessels (using the *erg/flk1* probe(McGary *et al.* 2010)) in *Xenopus laevis* embryos indicate the disruption of the vasculature in the

intersomitic (ISV, yellow arrows) and posterior cardinal veins (PCV, black arrows) caused by the presence of human and animal antihelminthics and broad-spectrum pesticides as compared to the DMSO control. Insets show growth profiles for yeast strains with humanized β-tubulin TUBB4 (orange) and TUBB8 (green) compared to wild-type (blue) when grown in the presence of each compound. Mean +/- standard deviation indicated by solid lines and shaded boundaries, respectively.

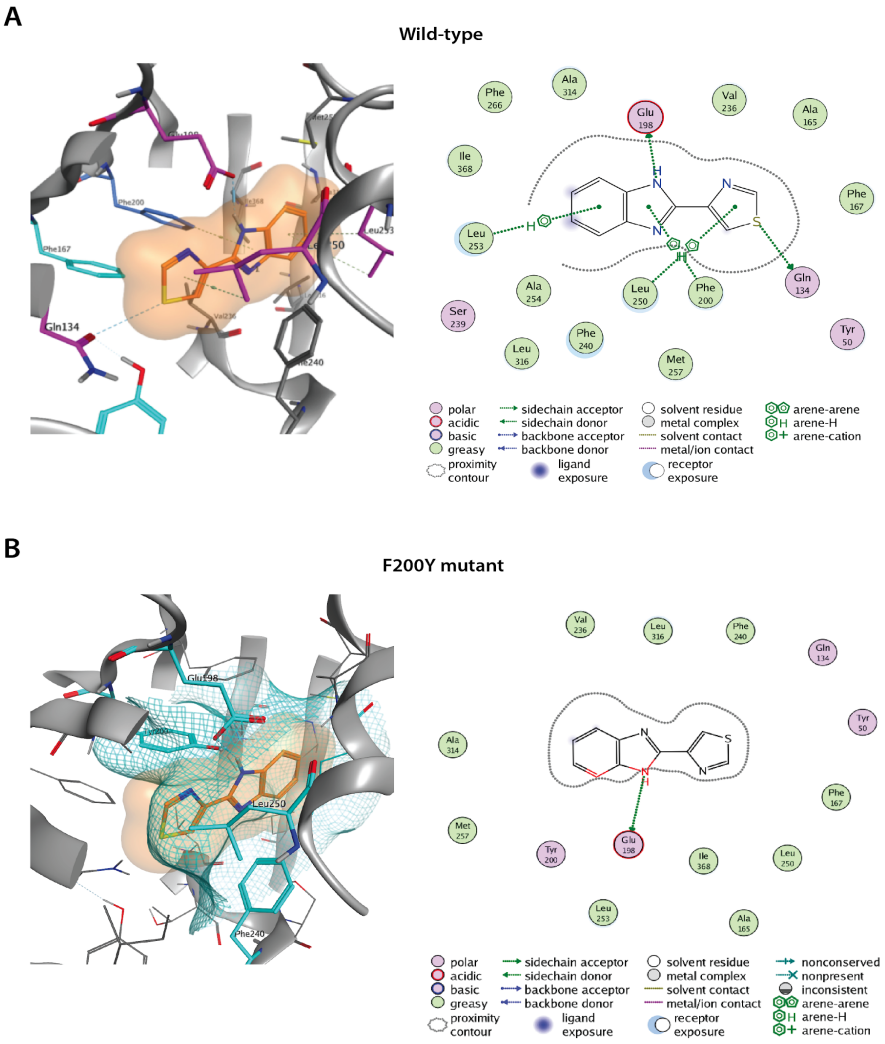


Figure 4.6.S1. Homology modeling and *in silico* docking studies predict the TBZ binding site in the fungal β-tubulin *NDA3* structure.

TBZ is well accommodated in wild-type *NDA3*'s predicted binding pocket (**A**) as opposed to its F200Y mutant (**B**). 3D structures and 2D contact maps shown on the left and right respectively indicate the steric clashes TBZ faces in the F200Y binding pocket. Cyan meshes (in 3D structures) and red highlights on the ligand (in 2D contact maps) indicate steric clashes in the binding pocket.

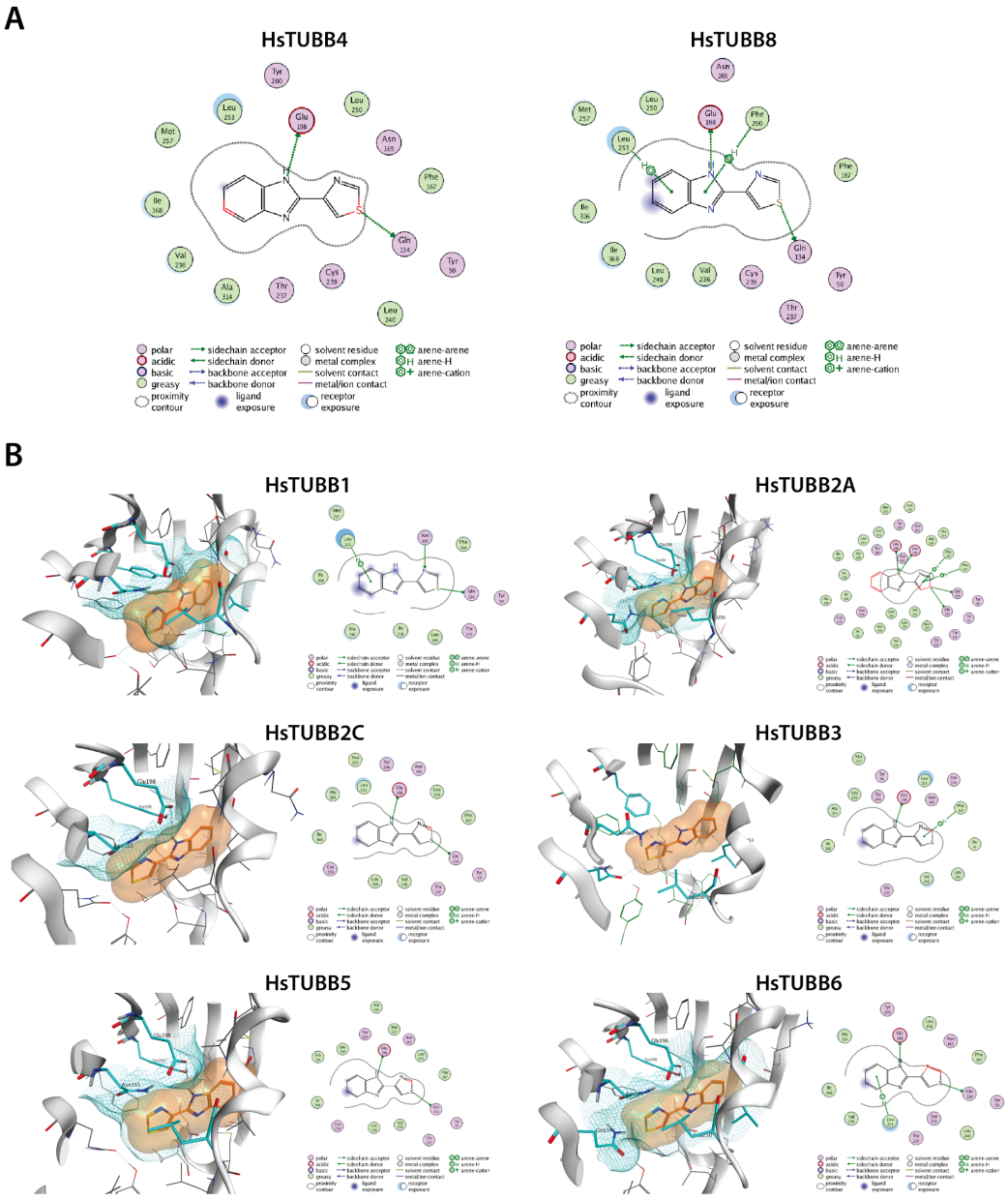


Figure 4.6.S2. Only TUBB8 favorably binds TBZ among the 9 human β-tubulins.

(**A**) 2D contact maps highlight ligand interactions between TBZ and TUBB4 (left) or TUBB8 (right). TBZ forms polar contacts with residues Q134, E198, F200, and L253 in

TUBB8. In TUBB4, reorientation of the proposed binding pocket is observed. Substantial steric clashing is shown in red on TBZ. (**B**) *In silico* docking of TBZ into other human β-tubulin homology models suggests substantial steric clashes due to unfavorable binding pockets among 8/9 human β-tubulin isotypes. Cyan meshes (in 3D structures) and red highlights on the ligand (in 2D contact maps) indicate steric clashes in the binding pocket.

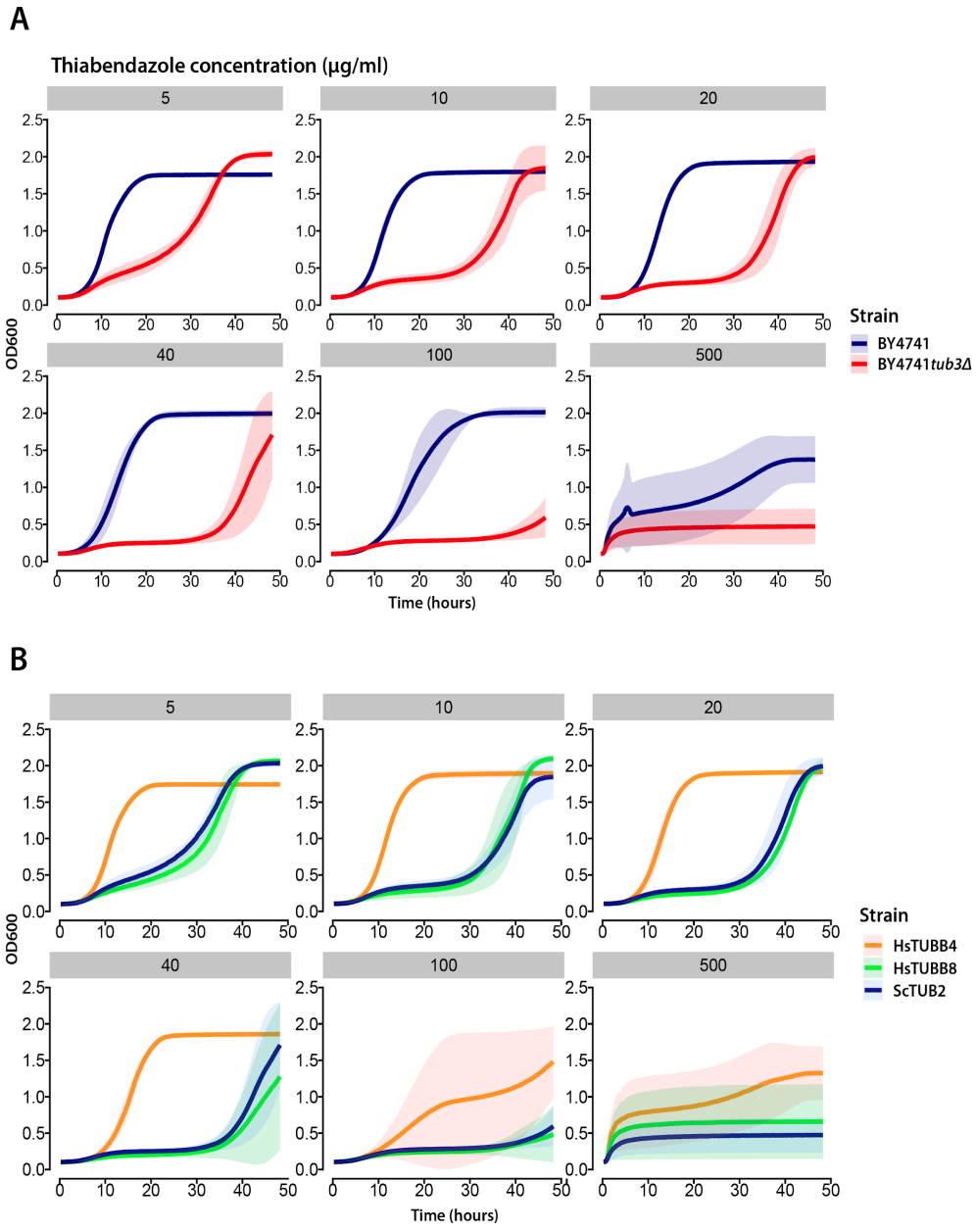


Figure 4.6.S3. Yeast strains with modified β-tubulin are differentially sensitive to TBZ.

(**A**) Deletion of yeast α-tubulin *TUB3* makes *Saccharomyces cerevisiae* (Baker's yeast) sensitive to TBZ. (**B**) Growth profiles of *tub3Δ* yeast strains with wild-type *TUB2* (blue),

humanized β-tubulins TUBB4 (orange), or TUBB8 (green) in increasing concentrations of TBZ show differential sensitivities to the drug.

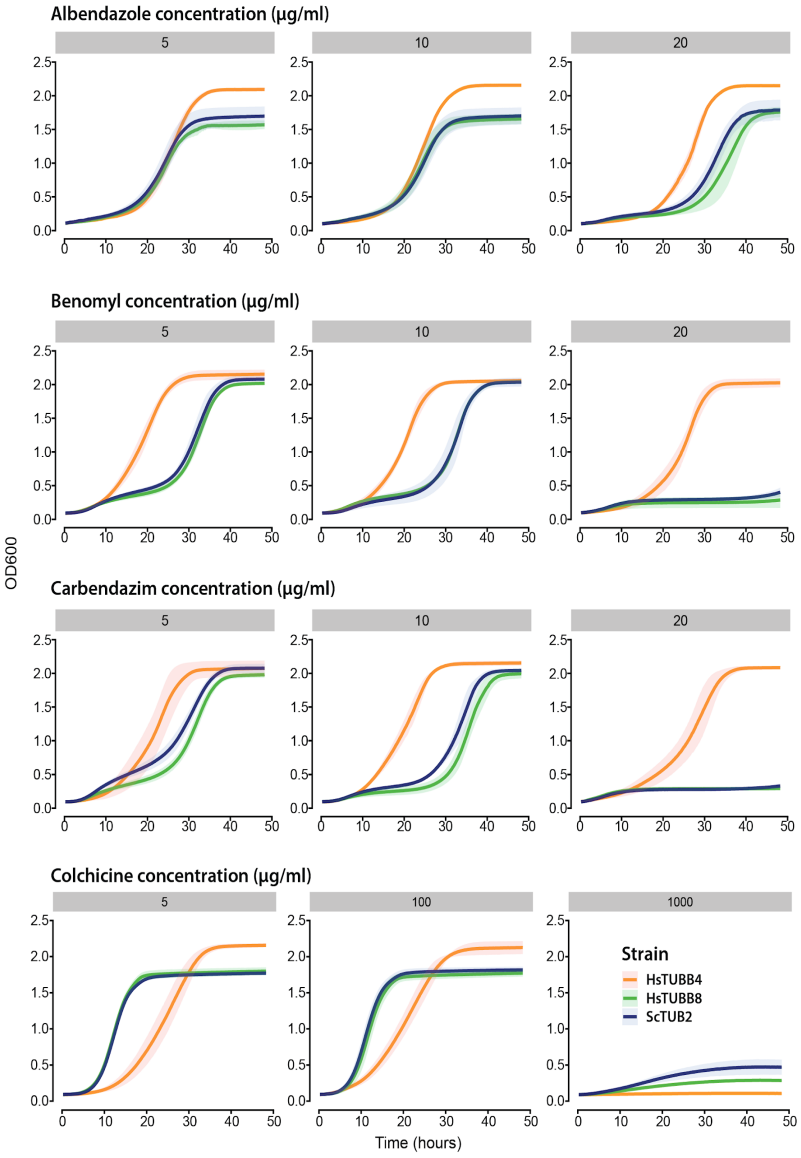


Figure 4.6.S4. Growth profiles of benzimidazole treated yeast strains.

Plots depict varying doses of albendazole, benomyl, carbendazim, and colchicine. Our data show that colchicine is a pan-isotype inhibitor whereas albendazole, benomyl, and carbendazim show some degree of specificity for TUBB8.

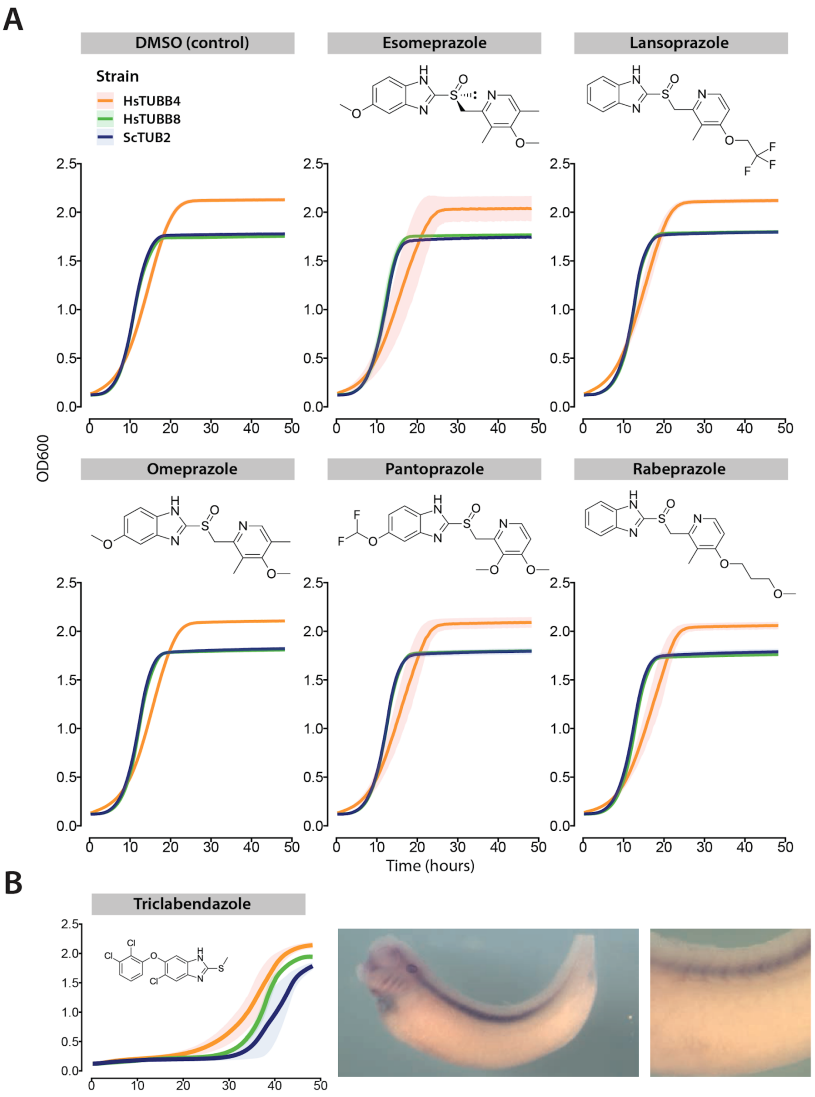


Figure 4.6.S5. Proton-pump inhibitors do not elicit growth defects in humanized strains.

(**A**) Yeast strains harboring the beta-tubulin gene *TUB2* (blue), TUBB4 (orange), and TUBB8 (green) are not inhibited by proton pump inhibitors (drug conc. 40 µg/ml). (**B**) Triclabendazole is a pan-isotype β-tubulin inhibitor inhibiting both wild-type and

humanized yeast strains (Left) and lethal to developing *Xenopus laevis* embryos at stage 38. (Right).

4.7. TABLES

Since the tables are very large, they are provided in the supplementary materials. Online versions are also available with this link.

Table 4.7.1. β-tubulin mediated resistance mutations seen in species curated from literature.

Table 4.7.2. Yeast site finder statistics for wild-type and F200Y.

Table 4.7.3. Docking free energy scores across β-tubulin isotypes.

Table 4.7.4. Yeast wt β-tubulin induced fit TBZ docking scores.

Table 4.7.5. Benzimidazole chemical features.

Chapter 5: Conclusions and Future Directions[4]

The central theme throughout this thesis is understanding the divergence in core biological systems over vast timescales using cross-species gene swaps in Baker's yeast. Chapter 2 systematically surveys the ability of orthologous (and paralogous) human genes in expanded gene families to functionally complement their yeast counterparts. Besides providing direct tests of function complementarity, it also reveals the extents to which ancestral roles are retained and/or distributed across gene families and the principles that underlie them. These assays with when combined with previous humanization efforts now include nearly all assayable essential yeast genes. Chapter 3 addresses functional divergence in gene families constituting a core eukaryotic module, the cystoskeleton. Expanding the scope of complementation assays in yeast beyond essential roles, these humanization experiments uncover the divergence of ancestral function in context-dependent and conditionally essential roles. Chapter 4 marks the intersection of multiple research threads including phenologs, cross-species gene swaps, and model organisms to discover the mechanism of vascular disruption underlying a class of widely used antifungal compounds. In Chapter 5, I discuss the directions that I anticipate being most interesting and present some preliminary work in humanizing entire yeast systems and their utility to understanding human health and disease.

[4]Most of the sterol pathway replacements in yeast were performed by Sophie Curie. The mass-spectrometry experiments of the humanized strains were performed by Daniel Boutz. The mu-opioid receptor signaling assays in yeast were performed by Colleen Mulvihill.

5.1. THE FUTURE OF HUMANIZED YEAST

Human gene complementation assays have been around for over 30 years now. Humanizing yeast genes is but a starting point. These assays are not new, however their utility to human biology (especially when paired with high-throughput yeast technologies) is. While this thesis predominantly uses humanized yeast to study divergence across evolutionarily conserved biological systems, its application to human functional genomics and systems biology is only in its infancy.

It is remarkable that on average roughly half of the human orthologs assayed are replaceable in yeast. This observation marks as a starting point for many research directions all worthy of pursuing. It begs the question of why the remaining genes were unable to. There may be several reasons for this. Most complementation assays in this thesis particularly assay essential roles of the yeast genes being queried. These assays while simple, may often fail to test the ability of human genes to rescue context-dependent and/or conditionally essential roles. For example, in Chapter 3 we found that despite only 4 human septins being replaceable, but all of them complemented the mating and meiotic roles of their yeast ortholog. Such findings are important as they point to fractional complementation of specific roles. Therefore, to exhaustively test replaceability, it is crucial to expand the scope of humanization assays in yeast to include non-essential roles as well. Another important caveat of humanization is recognizing the biological and technical limits of the available complementation assays. For example, sporulation-based assays inherently have an added layer of stringency in that they require human genes to appropriately function through meiosis while temperature sensitive assays only assay complementation in mitotic growth. Technically, there still remain

significant challenges in constructing robust reagent strains, curating accurate human clones and the cost of performing large-scale humanization in yeast. I foresee these challenges being overcome as the suite of yeast genome engineering tools continues to expand and the cost of large-scale DNA synthesis continues to drop.

5.2. BUILDING ENTIRE HUMAN SYSTEMS IN YEAST

Possessing a census of swappable genes in yeast now provides a global view of the human processes that can be studied in yeast. The first large-scale humanization assays provide reagent strains to study human genes individually. However, it is important to understand how multiple human genes interface with each other to execute their molecular and cellular roles. In the next phase, I anticipate multi-gene humanizations becoming common to study gene-gene interactions and epistasis. The efficiency of CRISPR-Cas9 based genome editing tools in yeast means that these gene replacements can now be multiplexed and even be performed iteratively. Perhaps it is not an unachievable goal to humanize every replaceable yeast gene natively under the influence of their respective loci. The advent of DNA synthesis technologies allows for testing human genes that were previously hard to assay and additionally enables large-scale testing of allelic variants for pathogenicity.

In recent years, we (Laurent 2016; Kachroo *et al.* 2017; Curie 2019) and others (Agmon *et al.* 2020) have pursued this direction. For example, we have systematically humanized ~80% of the yeast sterol biosynthesis pathway natively using CRISPR-Cas9 (**Fig. 5.6.1**). Additionally, by performing these replacements iteratively and in

combinations, these multi-gene humanized strains offer reagents to directly study feedback, epistasis and pathogenic allelic variants in the context of in the background of a full human module. I anticipate that such efforts building entire human systems in yeast becoming a fertile area of research in the years to come.

5.3. NON-ORTHOLOGOUS GENE SWAPS AND SURROGATE HUMAN SYSTEMS

Replacements need not be restricted only to orthologous genes. Genes (and even systems) may be replaceable without being orthologous. While I predict these instances to be few and far between, they may still yet prove useful. One example was the replacement of the yeast ergosterol biosynthesis enzyme *ERG8* by the human *PMVK* gene. Previously, a study found that the human *SEC61A1* was able to complement yeast strains lacking *RFT1* (Hamza *et al.* 2015). Recently, Luongo *et al.* found that the human gene SLC25A51 complements *NDT1* and *NDT2* deficient yeast strains by functioning as a mitochondrial NAD+ transporter (Luongo *et al.* 2020). With the widespread availability of large-scale yeast mutant libraries and human clones, such non-orthologous gene replacement warrants further exploration.

Perhaps our most exciting of non-orthologous replacements yet has been engineering cholesterol production in yeast. Zymosterol is the last common intermediate sterol between yeast and humans. Yeast and humans produce ergosterol and cholesterol as their terminal sterols (**Fig. 5.6.2A**). Ergosterol is a non-essential metabolite but strains lacking two or more genes in this pathway tend to elicit growth defects. We found that this pathway was entirely replaceable. Interestingly, strains with the partial pathway

showed strong fitness defects which could be rescued significantly when completely humanized (**Fig. 5.6.2B**). On assaying the protein expression levels, we find that these strains upregulate genes upstream of the pathway pointing to feedback and regulation (**Fig. 5.6.2C**). It begs the question of whether the upregulation of yeast enzymes at the feedback points is a modular property that can be restored when those enzymes are humanized too. Surprisingly, when engineered with appropriate enzyme expression, these strains produce cholesterol (**Fig. 5.6.3A & B**). It is interesting that when the pathway is natively replaced yeast strains do not produce cholesterol despite being viable offering interesting insights into replaceability beyond genes to metabolites. I believe that such engineering efforts in yeast will offer valuable eukaryotic surrogates to study human systems.

Perhaps the most promising direction with building non-orthologous human systems in yeast is repurposing these reagents as therapeutic platforms (as seen in Chapter 4) for human biology including drug discovery and modeling disease states. One direction we have pursued over the past several years is applying humanized yeast to understand G-Protein Coupled Receptor (GPCR) signaling. GPCRs play a crucial role in sensing external stimuli and constitute the largest class of membrane proteins in humans. About 34% of all FDA approved drugs target GPCRs (Hauser *et al.* 2018). Studying GPCR function in yeast proved challenging for a number of reasons including improper folding, expression, and different signaling environments. We found that modulating membrane sterol compositions in yeast through humanization was sufficient to enable human mu-opioid receptor signaling (**Fig 5.6.3C & D**). This is an exciting result because it not only offers orthogonal biosensor platforms to study receptor signaling, but

additionally examine the genetic tolerance in the receptors, discover ligands and small molecules capable of triggering orphan receptors whose mode of action is unknown. With the fast pace of large-scale pathway engineering in yeast, I believe that these studies will become more common and insightful for human health.

More broadly, non-orthologous gene swaps present opportunities to test hypotheses akin to The Ship of Theseus. As previously described, with the plummeting cost of DNA synthesis and advances in artificial chromosome assembly, extending humanization to include non-coding and regulatory regions is an achievable goal and would massively advance the field.

5.4. SACCHAROMYCES SAPIENS AND PERSONALIZED MEDICINE

The most exciting of directions for humanized yeast involve the ability to now personalize strains with an individual's allelic variant(s) either singly or in combination with interacting genes to better understand epistasis, pathogenicity and gene-drug interactions. It has been well established for a few years now that pathogenic variants and disease alleles can be efficiently discerned just by simple growth assays in yeast (Kachroo *et al.* 2015; Hamza *et al.* 2015, 2020; Sun *et al.* 2016, 2018; Yang *et al.* 2017). Humanizing libraries of gene variants *en masse* through deep mutational scanning (Fowler and Fields 2014; Fowler *et al.* 2014; Amorosi *et al.* 2021) coupled with functional screens in the context of partially or entirely humanized modules will significantly advance our understanding of genetic variance in populations and even offer insights for personalized therapies.

5.5. EVOLUTIONARY SYSTEMS BIOLOGY AND THE FUTURE OF MODEL ORGANISMS

Humanization of yeast is a lucid and effective approach to understanding and characterizing human biology. However, zooming out, cross-species gene swaps offer the opportunity to test multiple evolutionary questions. The obvious one that goes beyond humanization is the extent to which the trends and rules underlying replaceability are translatable to other species. When and how do orthologous genes lose the ability to work between species? Recent work from our lab has shown that yeast processes can be "plantized" and even "bacterialized" demonstrating that the rules of replaceability do indeed transcend domains (Kachroo *et al.* 2017). The extent of functional plasticity in core conserved biological machinery across the domains of life still remains to be seen. All gene swaps discussed up to this point have been performed in yeast, testing the orthologous gene of interest to function in a yeast cellular environment. An understudied aspect of complementation is the reciprocity of gene swaps. With mammalian cell culture and tissue engineering techniques advancing at a quick pace, perhaps yeastizing, plantizing and bacterializing human cells and/or other model organisms are all interesting and conceivable avenues worth pursuing. While cross-species gene swaps are strong tests of functional divergence, extending such approaches beyond genes to non-coding regions, chromosomes, cells, tissues offer far greater value especially when coupled with systems biology. The comparative nature of omics and systems biology along with the ever-expanding trove of phenotyping data available in model organisms altogether form the basis of contemporary biomedical research moving forward.

Finally, contemporary research is soaring to greater heights every day and at an unprecedented pace. This is largely due to the fact that science now lives in the data-driven age wherein multiple groups are parallelly generating a plethora of datasets from their experiments. Therefore, addressing challenges in data curation and consolidation will be key in devising and guiding future high-throughput experimental studies. The Saccharomyces Genome Database (SGD) has been at the forefront of database maintenance and management ever since its inception some 30 years ago including curating human complementation assays in yeast (Cherry *et al.* 1998, 2012). Recently, there has been a joint effort involving several model organism databases to consolidate functional gene annotations across multiple species (including model organisms) by The Alliance of Genome Resources (Agapite *et al.* 2020). Databases have been around for a couple of decades and it is crucial for us to recognize the roles that they play in shaping and guiding research. Therefore, in the era of big-data and multi-omics, model organisms are arguably more important than they have ever been. While there are exciting and promising efforts to shift large-scale genetic and biochemical studies into human cells, it is important to appreciate the roles model organisms have played in advancing our understanding of human health and disease over the decades. So, it is imperative that these databases be actively supported, and future curation efforts consider effective strategies to integrate and synthesize functional annotations across multiple species.

5.6. FIGURES

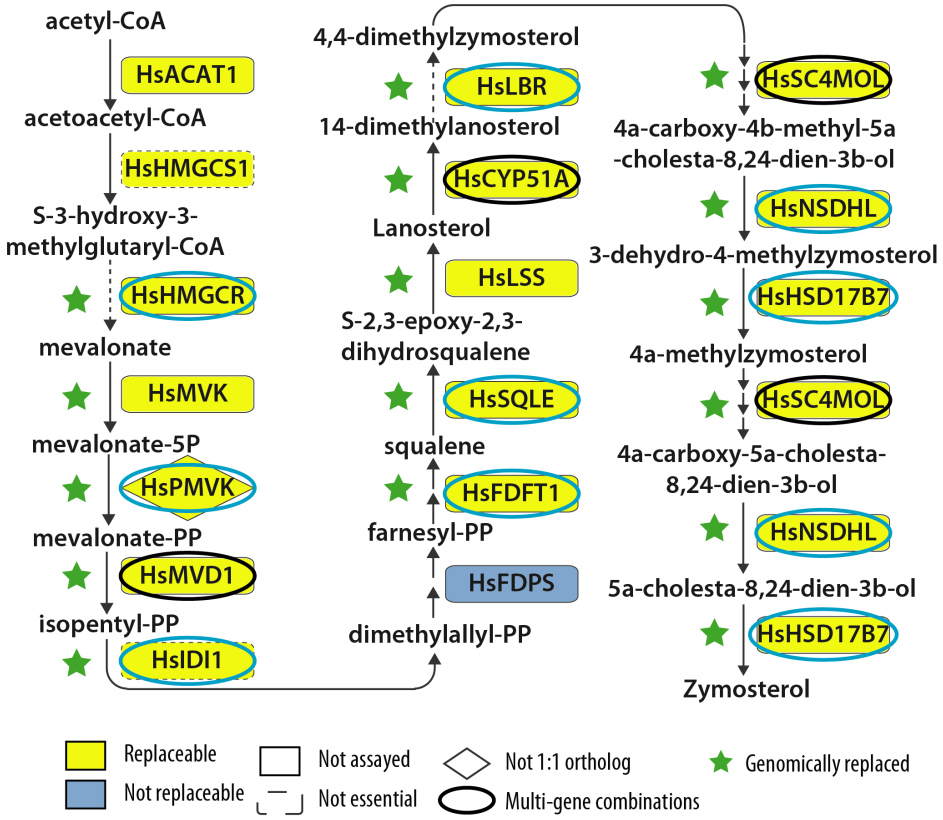


Figure 5.6.1. Entirely humanizing sterol biosynthesis in yeast.

The 19 enzymes from acetyl-CoA to zymosterol are conserved between humans and yeast. Majority of the pathway is replaceable (yellow). Using CRISPR-Cas9 genome editing 16 steps of yeast sterol biosynthesis have been genomically replaced with their human counterparts (Green stars). Further, using genome editing iteratively several multi-gene combinations can also be generated (colored ovals).

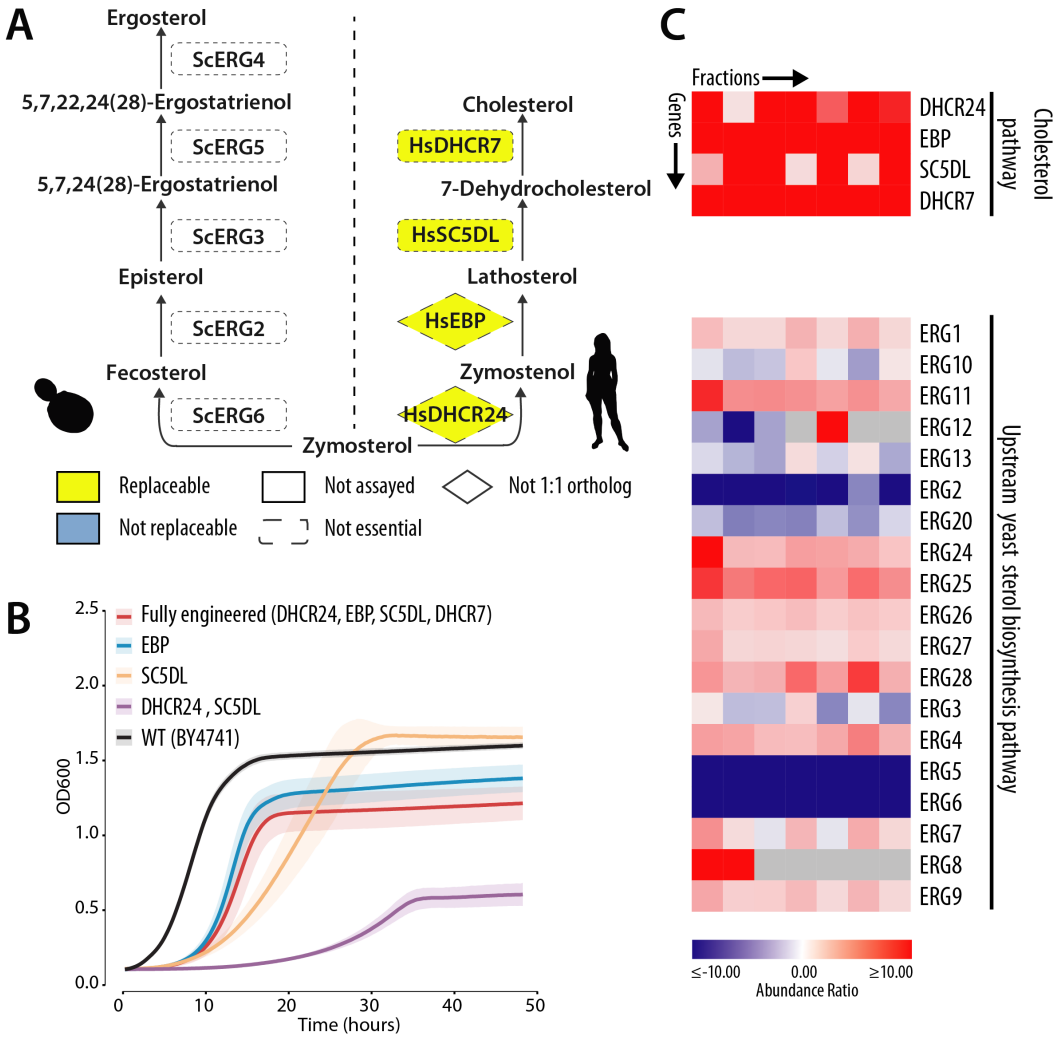


Figure 5.6.2. Humanizing yeast genes downstream of zymosterol.

(A) Yeast use 5 enzymes downstream of zymosterol to make ergosterol while humans use 4 to make cholesterol. The pathway was humanized in place their corresponding yeast genes i.e., DHCR24 in place of *ERG6* and so on. (B) Yeast growth profiles containing combinations of the cholesterol biosynthesis pathway. Singly humanized strains with EBP (blue) and SC5DL (yellow) have longer lag phases when compared to

wild-type (black). Doubly humanized strains with SC5DL and DHCR24 have strong growth defects (purple). Interestingly, despite having slower lag phase, humanizing the entire cholesterol biosynthesis pathway (red) restores doubling time close to wild-type. Lines and shades indicate mean and standard deviation respectively (n=3). (C) Proteomics of humanized yeast strains pathway uncovers feedback upstream of zymosterol. Heatmap depicting ratio of protein abundances (rows) in sterol biosynthesis across fractionation experiments (columns). Expression of human enzymes upregulates genes upstream of zymosterol. Particularly, enzymes *ERG11*, *ERG24*, *ERG25*, *ERG26*, *ERG27* are upregulated.

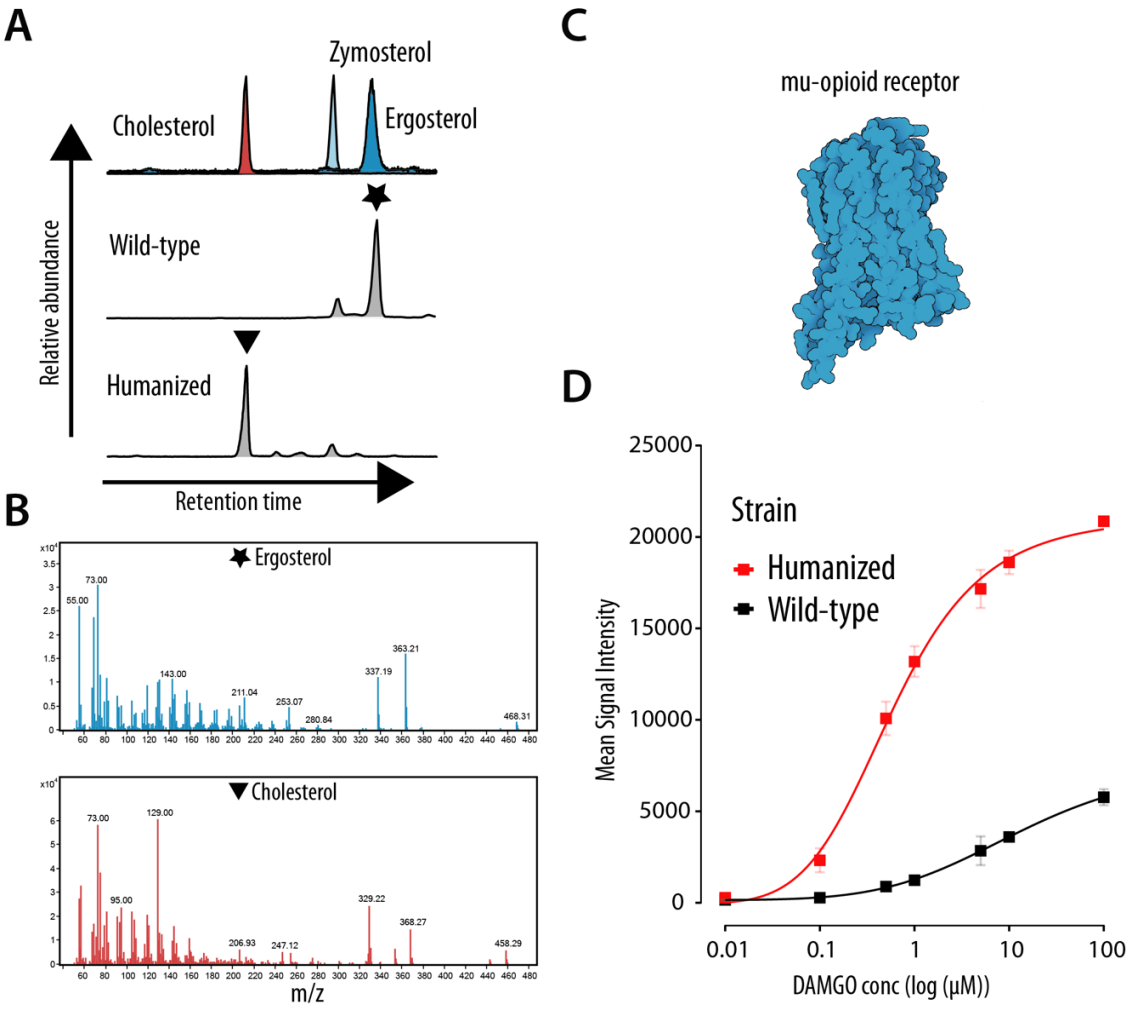


Figure 5.6.3. Repurposing humanized yeast as a platform to study GPCR signaling.

(A) Yeast sterol profiles analyzed by gas-chromatography mass-spectrometry (GC-MS). Top row indicates standards cholesterol (red), zymosterol (light blue), and ergosterol (blue). Wild-type strains (middle row) predominantly produce ergosterol while humanized strains (bottom row) produce cholesterol. (B) Mass spectra of ergosterol and cholesterol extracted from the labeled peaks from wild-type (star) and humanized strains (triangle) respectively. (C) Cartoon of *mus musculus* mu-opioid receptor *oprm1* (PDB ID-

5C1M). (D) Signaling of human mu-opioid receptor with its peptide ligand DAMGO in wild-type (black) and humanized (red) sterol biosynthesis backgrounds. EC50 for the humanized strain (0.54 μM) is 15 times lower when compared to wild-type (7.96 μM).

APPENDICES

Appendix A: Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9[5]

One of the main tools I have used over the course of my graduate work has been CRISPR-Cas9 which in recent years has become ubiquitous with synthetic biology and genome engineering especially in Baker's yeast. This project started to gain momentum right at the end of my first year of graduate school. Systematically screening the replaceability of hundreds of human genes in yeast warranted powerful tools to efficiently edit yeast genomes in a scalable fashion. Here we describe a one-step, scarless and marker-free method that was developed between the Gollihar, Ellington, and Marcotte labs to swap genes/ORFs in yeast. The most interesting aspect of this approach was its robustness to natively edit yeast genomes in a single step. Further, this method opens up avenues to engineer combinatorial and iterative multi-gene swaps to study epistasis via gene-gene interactions and entire modules in yeast.

A.1. ABSTRACT

Genome modification in budding yeast has been extremely successful largely due to its highly efficient homology-directed DNA repair machinery. Several methods for

[5]This chapter was published as Akhmetov, A., Laurent, J.M., Gollihar, J., Gardner, E.C., Garge, R.K., Ellington, A.D., Kachroo, A.H., and Marcotte, E.M. (2018). Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. *Bio-Protocol* 8. My contributions were performing the experiments and helping with the writing of the manuscript.

modifying the yeast genome have previously been described, many of them involving at least two-steps: insertion of a selectable marker and substitution of that marker for the intended modification. Here, we describe a CRISPR-Cas9 mediated genome editing protocol for modifying any yeast gene of interest (either essential or nonessential) in a single-step transformation without any selectable marker. In this system, the Cas9 nuclease creates a double-strand break at the locus of choice, which is typically lethal in yeast cells regardless of the essentiality of the targeted locus due to inefficient non-homologous end-joining repair. This lethality results in efficient repair via homologous recombination using a repair template derived from PCR. In cases involving essential genes, the necessity of editing the genomic lesion with a functional allele serves as an additional layer of selection. As a motivating example, we describe the use of this strategy in the replacement of HEM2, an essential yeast gene, with its corresponding human ortholog ALAD.

A.2. INTRODUCTION

Saccharomyces cerevisiae (Baker's yeast) has a long history as a genetically tractable organism, and there is an array of methodologies to manipulate the yeast genome. However, until recently it has been necessary to apply selection to isolate clones possessing the desired genetic alteration (Kearse *et al.* 2012; DiCarlo *et al.* 2013; Lee *et al.* 2015; Kachroo *et al.* 2017). In cases where arbitrary, scar-less editing of the genome is desired, the solution is typically a two-step process: First a selectable cassette (containing the URA3 marker, for example), flanked by homology arms targeting the

region of interest, and sometimes containing nuclease targeting sites (*i.e.*, I-SceI sites) to aid in the removal of the cassette at the later stage, is knocked in via homologous recombination (HR). The small subpopulation of successful integrants are isolated by selecting for the cassette. Second, the marker is eliminated through highly efficient sequence specific methods such as site-specific recombination or endonuclease cleavage (I-SceI) to generate the desired form of the edited genomic locus. Two steps are necessary because no method was available which is both scar-less and efficient enough such that no selection is required.

The development of CRISPR/Cas9 technology in yeast has eliminated the need for this two-step process. Cas9 efficiently creates double-stranded breaks (DSBs) in yeast DNA at virtually any arbitrary locus provided a PAM sequence is proximal to the desired cut site. When an appropriate repair template is provided, these DSBs are repaired through the endogenous HR system of yeast. Cas9 directed to the desired genomic locus via the guide RNA sequence creates double-stranded break (DSB) in the genome. The CRISPR target site is retained in cells which fail to repair the target site as expected, which allows Cas9 to repeatedly cleave the same region until HR-mediated editing takes place. Rarely, non-homologous end-joining (NHEJ) can generate mutations which block Cas9 cleavage despite failing to incorporate the expected genomic alterations. More commonly, cells simply succumb to the stress of repeated Cas9-induced genomic cleavages. In an appropriately conducted experiment, the majority of the surviving population tends to be cells which have lost their CRISPR target site by incorporating the desired genomic alteration via HR. Cas9 thus acts as a counter-selection acting directly on genomic sequence, rather than its phenotypic manifestations.

Here, we use an approach developed by Dueber and colleagues (Lee *et al*., 2015) to rapidly generate single, self-contained plasmids that express both the Cas9 nuclease and guide RNA required for targeting a desired locus. These plasmids, when co-transformed with an appropriate repair template provided as a linear PCR product, allow efficient, precise, single-step replacement of any arbitrary yeast gene with an introduced sequence of interest. Only selection for the Cas9 and gRNA-expressing plasmid is required, which tends to select for correct genomic modification by proxy due to efficiency of targeting and repair. This strategy was used extensively in our ortholog complementation research (Kachroo *et al*., 2017) to rapidly humanize, bacterialize and plantize many essential yeast genes. A CRISPR based approach is uniquely suited to this case, because it strongly encourages HR with functional alleles. False positives, arising from CRISPR sites being mutated by NHEJ without incorporation of a new allele, are minimal because they are often not viable. Additionally, disruption of the target gene's function is brief, eliminating the need for constructing and maintaining a complementing plasmid to sustain yeast through an otherwise lengthy engineering process. Further, given that CRISPR selects against sequence regardless of function, it is still possible and practical to alter non-essential genes (or even non-genic regions) with this technique; indeed, we have reported successful humanization of the non-essential yeast gene HEM14 with this method (Kachroo *et al*., 2017) and we have used this system to incorporate site-directed changes in proteins with high efficiency.

A.3. MATERIALS AND METHODS

1. Reagents

1. Pipette tips (Mettler-Toledo, catalog number: 17005872, 17005874, 17007089)
2. 96-well plate (VWR, catalog number: 82006-636)
3. Yeast (BY4741)
4. MoClo Yeast Toolkit (YTK, Addgene kit, catalog number: 1000000061). Toolkit includes plasmids pYTK050, pYTK003, pYTK072, pYTK083, pYTK036, pYTK008, pYTK047, pYTK073, pYTK074, pYTK081 and pYTK084.
5. NEB 5-alpha Competent *E. coli* (NEB, catalog number: C2987)
6. T7 ligase (NEB, catalog number: M0318S)
7. T4 ligase buffer (NEB, catalog number: B0202S)
8. Restriction enzymes *Bsa*I (NEB, catalog number: R0535S) and *Bsm*BI (NEB, catalog number: R0580S)
9. PCR template for the sequence which will replace the target gene (*e.g.*, cDNA, plasmid-based clone, *etc*.)

   Note: For demonstration purposes, this protocol will assume replacement of S. cerevisiae HEM2 with its human ortholog ALAD.
10. High-fidelity DNA polymerase for repair template PCR, such as KAPA HiFi (Kapa Biosystems, catalog number: KK2601)
11. Agarose (Thermo Fisher Scientific, catalog number: 16500500)
12. DNA stain (Thermo Fisher Scientific, catalog number: S33102)
13. Zymo DNA Clean&Concentrator-25 kit (Zymo Research, catalog number: D4005)

14. Zymo EZ yeast transformation II kit (Zymo Research, catalog number: T2001)
15. Optional: 100 mM lithium acetate can be used in place of EZ 1 solution from the EZ competent yeast cell kit. Lithium acetate can be obtained from Sigma-Aldrich, catalog number: L6883. See Recipes.
16. Agar (Serva, catalog number: 11396)
17. LB Broth, Lennox (Fisher Scientific, catalog number: DF0402-07-0)
18. LB plates with antibiotic selection (see Recipes)
    a. Ampicillin (Sigma-Aldrich, catalog number: 10835242001)
    b. Spectinomycin (Sigma-Aldrich, catalog number: PHR1426)
19. Chloramphenicol (Sigma-Aldrich, catalog number: C0378)
20. YPD powder (Fisher Scientific, catalog number: DF0428-17-5)
21. Yeast nitrogen base without amino acids (Fisher Scientific, catalog number: DF0919-15-3)
22. Ammonium sulfate (Sigma-Aldrich, catalog number: A4418)
23. Dextrose (Avantor, catalog number: 1919)
24. SC-Ura dropout powder (Sigma-Aldrich, catalog number: Y1501)
25. Liquid YPD medium (see Recipes)
26. YPD and SD-Ura agar plates (see Recipes)
27. Optional: 5-fluoroorotic acid (Sigma-Aldrich, catalog number: F5013), if counter-selection will be used (see section titled “Curing of the CRISPR plasmid”).
28. Zymolyase (MP Biomedicals, catalog number: 320921)
29. D-Sorbitol (Sigma-Aldrich, catalog number: S3889)
30. Accuprime Pfx (ThermoFisher Scientific, catalog number: 12344024)

31. Zymolyase solution (see Recipes)

2. Equipment

1. 12-channel pipette (Mettler-Toledo, catalog number: 17013810)
2. Thermocycler (Bio-Rad, catalog number: 1861096)
3. Standard gel electrophoresis tank and accessories (Bio-Rad, catalog number: 1640302)

3. Software

1. Geneious v8.0 (Kearse *et al.* 2012) or higher, to design gRNA and repair template (replacement gene). Other gRNA design software can be used as well, such as E-CRISP (Heigwer *et al.* 2014).
2. BLAT (Kent 2002).

A.4. PROCEDURE

1. Preparation of CRISPR plasmid (for a diagrammatic overview of the cloning process, see **Figure A.8.2**). Design two guide RNA (gRNA) sequences targeting the open reading frame (ORF) for the yeast gene to be replaced using Geneious, or a similar tool such as E-CRISP (Heigwer *et al.* 2014). gRNA sequences can often have low activity in practice, despite being predicted to be highly efficiencient by software tools. In order to minimize setbacks due to a gRNA which turns out to function poorly, we advise designing multiple gRNAs from the outset, and taking them through the cloning steps in parallel, up to and including

the construction of the CRISPR plasmids. Both plasmids should then be tested for their ability to target the yeast genome and kill cells (described in later steps) to empirically determine and confirm their activity. We have not noticed a strong effect of the location of the gRNA within the ORF. During homologous repair, DNA can be resected up to several kilobases from the break site(Mimitou and Symington 2009; Chen *et al.* 2011), so the gRNA need not be very close to either terminus of the ORF. It is however important to select a gRNA such that the target site is not present in the eliminated by the replacement (*i.e*., the gRNA should target the yeast ORF, but not the replacement gene). Example: For targeting HEM2, the sequences GGATTATCGGAGATGAATAG ('sg1') and CCTGGTACCAAGGATCCAGT ('sg2') were predicted to have high activity (see **Figure A.8.1**). Order forward and reverse oligonucleotides with the gRNA sequence and Golden Gate compatible overlaps: Forward oligo consists of the 5' insert GACTTT followed by the 20 bp guide sequence specific to the target gene. Example forward oligo for HEM2 sg1 (underline indicates 5' Golden Gate overhang): GACTTTGGATTATCGGAGATGAATAG. Reverse oligo consists of the 3' insert AAAC, followed by the reverse complement of the 20 bp guide sequence, and a 5' insert TT. Example reverse oligo for HEM2 sg1 (underline indicates 3' Golden Gate overhang): AAACCTATTCATCTCCGATAATCCAA. Mix forward and reverse oligos (50 μM each) for each gRNA in a total volume of 20 μl and anneal with each other using a thermocycler with the program below. It is unnecessary to phosphorylate the insert or vector. 95 °C for 5 min, 55 °C for 15 min, 25°C for 15 min.

2. First Golden Gate cloning reaction to transfer into shuttle vector: Set up cloning reaction with annealed oligos and pYTK050 (Table A.9.1). A 2:1 molar ratio of insert:plasmid is recommended for optimal Golden Gate cloning of linear DNA.
3. Transform the reaction into competent bacteria and plate with chloramphenicol selection (170 µg/ml). View colonies under UV light and pick the white colonies (those not showing GFP fluorescence), then grow in liquid culture and purify plasmid. The vectors used in Golden Gate reactions described in this protocol are all GFP-dropout vectors: They contain a GFP gene which will be silenced upon successful cloning. Therefore, GFP fluorescence indicates an invalid construct, while successful constructs will lose the GFP gene and thus resulting colonies will be white. Optionally, the plasmid can be sequenced to check for errors or mutations in the gRNA sequence, such as may occur during synthesis.
4. Second Golden Gate cloning reaction to create gRNA cassette plasmid: Set up cloning reaction which includes connector plasmids ConL1 and ConRE (Table A.9.2). For best efficiency, all plasmids should be present at the same molarity in plasmid based Golden Gate assemblies.
5. Transform the reaction into competent bacteria and plate with ampicillin selection (60 µg/ml). View colonies under UV light and pick the white colonies (those not showing GFP fluorescence), then grow in liquid culture and purify plasmid.
6. Third and final Golden Gate cloning reaction to construct the yeast-compatible, complete CRISPR plasmid: Set up Golden Gate cloning reaction with connector plasmid from previous step, and yeast –Ura backbone plasmid, and Cas9 plasmid (Table A.9.3).

7. Transform the reaction into competent bacteria and plate with kanamycin selection (50 μg/ml). View colonies under UV light and pick the white colonies (those not showing GFP fluorescence), then grow in liquid culture and purify plasmid.
8. The resulting construct is a self-contained CRISPR plasmid, which when transformed into yeast will cause double-stranded breaks (DSBs) at the locus determined by the gRNA sequence cloned into it. 500 ng of this will be used for each yeast transformation, so if multiple replacements are planned it is helpful to dilute the CRISPR plasmid to a standardized concentration for easier transformation set up later on.
9. Preparation of repair template DNA. Design the template DNA using Geneious or any other cloning software. Obtain the genomic sequence of the target yeast gene ('old gene'), and the coding sequence (CDS) of the replacing gene ('new gene'). The CDS should not contain introns. Create a gene model for the replaced locus by editing the sequence of the old gene so that it contains the new gene in the correct position (*i.e.*, the desired outcome of replacement).

   We find that replacement works best if the original yeast stop codon is left intact. Otherwise, modifying the new gene, for instance to codon optimize for yeast, has proven unnecessary.
10. Design template PCR primers which anneal to about 25 bp of the 5' and 3' ends of the new gene's CDS, and also the 5' and 3' UTR immediately adjacent to the ORF (the homology arms). **Figure A.8.3** shows an example of primer design for replacing the yeast HEM2 gene with its human ortholog ALAD. This process is

much easier using the gene model constructed in the previous step: The sequence covering the junction points between yeast genome and the new gene CDS can be used directly as primer sequence.The length of the region complementary to the new gene CDS is determined only by standard PCR efficiency concerns, such as melting temperature. This area will serve as a toehold for the first few cycles of the PCR. The length of the homology arms is critical for efficient replacement. We find that homologies of at least 70 bp are necessary (in which case the entire primer oligo will be about 90 bp long), and for some genes, 170 bp homologies may be necessary. For even more difficult replacements, longer homology arms can be cloned separately, but we have found that homologies longer than 500 bp are unlikely to increase efficiency further.

11. Use template PCR primers to amplify a large amount of repair template DNA using a high-fidelity polymerase. We find that it is helpful to first conduct several test PCRs with different polymerases. Due to the particular design of the template primers, this PCR can sometimes run inefficiently or generate unwanted non-specific products. Different polymerases have different characteristics, and often a reaction which fails with one polymerase will run efficiently with another, rendering laborious PCR optimization unnecessary. At least 5 µg of template DNA is needed per yeast transformation, which can usually be obtained from a single 50 µl PCR. Difficult replacements can often be facilitated by using more (10 µg) template DNA, and if multiple transformations are to be performed the amount will also need to be scaled up accordingly. Often several PCRs are necessary to produce enough DNA. If very large amounts of template DNA are

needed, or an efficient PCR is difficult to set up, an alternative method is to clone the template sequence onto a plasmid, which can be amplified in bacteria with the template DNA excised using restriction enzymes.

12. Check the template PCR with agarose gel electrophoresis. As long as a sufficient amount of the correct template is produced, non-specific products do not necessarily constitute a problem for the replacement. Because the non-specific products usually lack appropriate homologies, they will not be efficiently integrated into the yeast genome. However, if significant amounts of them are present, they will cause over-estimation of template DNA during spectrophotometry-based quantification; thus the amount of template DNA used in the transformation would need to be adjusted accordingly. Alternatively, the PCR can be optimized to reduce non-specific products, or only the correct product can be quantified from the gel using a DNA ladder calibrated for quantity estimation.
13. Purify template PCR using the Zymo DNA Clean&Concentrator-25 kit. Elute in water. Ideally, the volume of DNA included in yeast transformation should be small, so as to not interfere with the transformation reagents. The elution volume should be adjusted accordingly so that the resulting concentration of DNA is not too low. In our experiments, we have found that eluting with 25 µl water will usually yield 400-800 ng/µl DNA, which is suitable for transformations.
14. Yeast transformation: Prepare competent yeast cells using the Zymo EZ competent yeast kit according to the kit instructions. The EZ 1 solution in this kit can be substituted with 100 mM lithium acetate without significant change in

transformation efficiency. The amounts given in the kit manual can be slightly modified: 2 ml yeast culture can be used to produce 100 µl of competent yeast, which is sufficient for two transformations, 50 µl each.

15. Set up a transformation reaction: Mix 50 µl competent yeast, 500 µl EZ 3 solution, 500 ng of CRISPR plasmid and 5 µg repair template DNA (up to 50 µl total volume). Incubate at 30 °C as directed by kit manual and plate on –Ura medium.
16. When using a new gRNA for the first time, gRNA efficiency can be estimated with a control transformation, which is performed as stated but without repair DNA. When the CRISPR plasmid is introduced without a repair template, it will repeatedly cleave the target locus, causing toxicity. Very few or no colonies are the ideal outcome, since this indicates highly efficient CRISPR cleavage and low background rate. Cells can survive the CRISPR plasmid uptake without repair DNA if the CRISPR activity is stochastically low (such as due to poor gRNA efficiency) or mutations at the CRISPR target locus can be tolerated (which produces false transformants even in presence of the repair template).
17. When colonies appear on the –Ura plates, collect up to 12 of them with a pipette tip and suspend in 50 µl water. These suspensions will be screened for confirmed replacements. Yeast suspensions can be stored at 4 °C and used to start new cultures for up to 2 weeks. Typically, colonies will appear on –Ura plates (**Figure A.8.4**) after 1-3 days. In some cases, the replacement will impose a significant fitness defect such that up to 6 days may be required for colonies to appear, but we have not encountered cases where colonies from a successful transformation

take longer than 6 days to grow. The uracil dropout medium will select against cells which failed to take up the CRISPR plasmid (which confers uracil prototrophy), but because the CRISPR plasmid is toxic to cells unless a successful replacement occurs (eliminating the CRISPR target locus) only cells which have a replaced locus are expected to survive. However, due to spontaneous hypoactivity of the CRISPR system, mutations in the CRISPR target locus (DiCarlo *et al.* 2013), and cells which manage to survive CRISPR-associated DSBs, there will be a background rate in the form of false transformant colonies which do not carry the correct genomic replacements. To save time, we recommend collecting several transformant colonies and screening them in parallel. To streamline this process (especially when several replacements are performed in parallel), pick colonies with pipette tips and manually attach them to a multichannel pipette. The multichannel pipette can then be used to suspend all 12 samples in one row of small PCR tubes or a 96-well plate.

18. Colony screening via PCR. Design confirmation PCR primers: Primer pairs should be selected such that the forward primer anneals to the yeast UTR while the reverse primer anneals only to the new gene CDS but not the old gene's ORF. Thus, the product should span the junction point between foreign sequence and native yeast genome. The yeast UTR primer should preferably not overlap the homology region. Ideally the product size should be small, about 300 bp, for a faster and more robust PCR. It is sufficient to check only the 5' junction point, since it is rare for integration to be proceed as expected at one end of the gene but introduce artifacts at the other. If desired, the absence of the yeast ORF can also

be tested by using a reverse primer which anneals to yeast ORF only. However, lack of product from such a primer pair is not sufficient to confirm a clone, since the reaction is liable to fail for unrelated reasons (such as poor lysis of cells).

19. Prepare lysates of harvested transformants: Mix 5 µl of each yeast suspension with 15 µl zymolyase solution.
20. Incubate lysates for 30 min at room temperature, then 15 min at 37 °C and 5 min at 95 °C.
21. Set up 20 µl colony PCRs with confirmation primers and using Accuprime Pfx as the polymerase. Use 1 µl of the lysate as template DNA. We find that other polymerases do not perform well due to impurities from the yeast lysates. Due to the impurities introduced by the lysate, the colony PCR may spontaneously fail, leading to false negatives. To ameliorate this problem, a positive control PCR can be performed for each lysate, which is identical to the confirmation PCR but uses primers complementary to an unrelated, unmodified locus in the genome. We use two primers targeting a 500 bp segment of the yeast ERG13 promoter for this purpose (Forward CGAACTGGATGAGATGGCCG and reverse CATGCTGCACCTTTTATAGTAATTTGGC)
22. Check the colony PCRs for product by agarose electrophoresis. Lysates from clones with the correct modifications should generate a product with the confirmation primers. Background false transformants (*e.g.*, mutants) will not produce a band. A PCR product from the confirmation primers is sufficient evidence of successful integration of the repair template. For further verification, the locus can be sequenced, but we have found that dramatic sequence artifacts

rarely occur in clones confirmed by PCR, the most common mutations are single-basepair substitutions or indels, which typically constitute a minority of confirmed clones. Lack of product from the confirmation primers is inconclusive per se. In such cases, it is worthwhile to consider additional evidence, such as whether the positive control PCR worked (if not, the lysis may have failed).

23. Confirmed clones can be propagated by starting a new culture from the original suspensions of yeast in water.
24. Curing of the CRISPR plasmid. Streak original water suspensions of confirmed clones on YPD. The CRISPR plasmid is low copy and can be spontaneously lost in absence of selection. Pick 10 colonies from the YPD plate and patch each one on YPD and SD-Ura plates. Incubate both plates and collect cells from patches which grew only on YPD but not on SD-Ura. Isolates which still carry the CRISPR plasmid will grow on uracil dropout medium, but those which have lost the plasmid will not. Typically, 3 days is sufficient to confirm lack of Ura prototrophy, but if slow growth on uracil dropout is suspected, incubation can be extended to up to 6 days to definitively confirm no growth on uracil dropout. The plasmid can also be cured by counterselecting on 5-fluoroorotic acid (FOA) plates (Boeke *et al.* 1987). However, there is a possibility that this FOA method will generate some colonies that are not cured of the plasmid but rather have acquired a mutation in the Ura marker (thus continuing to express the gRNA). Thus, FOA counterselection should not be used (as opposed to replicate patches on YPD and –Ura) if it is important to ensure curing of the plasmid, rather than simply abrogating Ura prototrophy. On the other hand, the FOA method can save time if

only loss of –Ura heterotrophy is desired, for instance to enable a subsequent transformation with a different Ura-selectable plasmid.

A.5. Data Analysis

The data analysis needs for this procedure are minimal. Most importantly, when using Geneious to design gRNA sequences, it is desirable to select gRNA sequences that have high predicted on-target activity (automatically calculated by Geneious). gRNA sequences with high predicted activity may have low actual activity, but they will be less likely to exhibit low activity than sequences with low predicted activity. The distance of the gRNA target site can be up to 1 kb away from either homology region without perceptible negative consequence, thus gRNAs should be selected primarily based on high activity rather than location (provided that they lie between the two homology arms).

A.6. Notes

1. We have found that even among gRNAs with high predicted activity, some will fail to induce double strand breaks with sufficient efficiency for editing. It is highly recommended that for each target locus, several gRNA are designed and tested in parallel, to ensure that at least one will be a sufficiently good DSB inducer for purposes of genome editing.

2. If a given gRNA exhibits significant off-target activity, the likely outcome is that off-target cleavage will kill most of the transformed yeast cells. Successful, efficient genome editing relies on lethality associated with DSBs at the target locus being rescued by HR (allowing efficient repair of the DSB) and abrogation of the gRNA target site (preventing further cleavage). In the event off-target activity, HR may likely not take place because no repair template with homology to the off-target site has been supplied, moreover the gRNA site will not be eliminated for the same reason. Further, the confirmation strategy we suggest is such that only repair at the correct locus will produce a positive result. However, it is nevertheless worthwhile to ensure that selected gRNA target sites do not occur at other locations in the genome, where cleavage is not intended. Although it is very unlikely for the combined 23 bp target sequence to appear multiple times in the yeast genome, we recommend confirming that candidate gRNA sites appear only in the target locus using a tool such as BLAT.
3. gRNA targets consist of a 20 bp sequence (which will also be included in sgRNA sequence and become part of the Cas9 complex) followed by a 3 bp PAM sequence (which takes the form of NGG for Cas9 described in this protocol). The PAM sequence does not become part of the gRNA, but it must be present in the target genome for Cas9 cleavage to occur. This can be verified by attempting to align the gRNA sequence to the sequence of the repair template – typically, CRISPR activity will be very low with more than 5 mismatching basepairs, although mismatches in the PAM and proximal to

the PAM appear to have more significance (Kuscu *et al.* 2014). When replacing with very similar sequences, such that it is difficult to find good gRNA sites unique to the target locus, one strategy that can be adopted is to introduce synonymous mutations in the repair template sequence which alter the PAM site or PAM-proximal nucleotides. Alternatively, recent research suggests that using shorter gRNA may increase specificity, since the 8-17 PAM-proximal nucleotides contribute disproportionately to CRISPR target recognition (Xu *et al.* 2017).

4. There is some variability in the yeast transformation step, and depending on how the competent cells were prepared, and how the transformation was performed. Most commonly, the number of resulting colonies will vary somewhat between transformations of identical strains with identical reagents, but usually this variation will be less than tenfold. When a transformation produces a fair number of colonies (at least 10) yet none of them are found to be correct clones upon screening, simply repeating the transformation is unlikely to improve results. The most straightforward avenues of increasing the number of correct clones are to increase the amount of repair template DNA, and to produce repair template DNA with longer homologies.
5. If no colonies appear after transformation, the reason may be low transformation efficiency. In this case, several troubleshooting steps can be taken (described in detail in the documentation of the Zymo EZ competent yeast kit). We have found the following to be effective: Thoroughly vortexing the mixture of competent cells and DNA. Longer incubation time for the

transformation (1.5 hours instead of the 45 minutes). Including more cells in the transformation. Competent cells seem to perform slightly better when frozen once (slowly in -80 °C) than freshly prepared cells.

6. When the CRISPR reagents and repair template are transformed into yeast cells, the resulting transforming colonies will be of three kinds with respect to the targeted locus: Correct transformants which bear the sequence of the repair template. False transformants which bear the original, unedited sequence. Mutants
7. In our experiments, we have found that the first two classes predominate unless mutants are specifically selected for. Even in the absence of a repair template, the majority of false transformants will not be mutants. Due to the efficient HR system of *S. cerevisiae*, if the conditions of the experiment are adequate then editing will take place at a very high rate. Thus, typically, the proportion between the first two of the three classes listed above will be such that the transformants are either mostly correct or all false. The third class, or mutants, we have found to be very rare in either case unless specifically selected for. As a consequence, it is rarely necessary to screen a very large number of colonies to determine whether an editing experiment has succeeded. However, it is desirable to collect several confirmed clones to minimize issues caused by artifacts, such as for instance mutant edited sequence caused by errors during PCR (with the reagents and protocols described in this text, we have found clones with mutant edited sequence to also be very rare).

8. Selecting yeast transformants with a single amino-acid dropout medium is normally a straightforward process, and colonies can be seen within 1-2 days of plating. However, occasionally the genome editing process itself, or the resulting edited sequence, can result in a growth defect in the resulting cells. Thus, if no colonies appear, incubating the plate for a longer period can produce colonies. In the most extreme case we observed, it took 6 days for colonies to appear on a uracil dropout medium, but several clones were later confirmed by PCR and sequencing; these clones consistently exhibited slow growth in subsequent culture on rich medium (YPD) as well.
9. Some combinations of target locus and repair template may lead to a mixture of large and small yeast colonies after transformation. If this occurs, generally it is best to screen an adequate number of colonies for each size class. It may be that the correct edits create much slower growing strains, thus the large colonies are false while the small ones have the desired edit. Conversely, if the desired sequence does not interfere with normal growth, but mutations arising from NHEJ do, then larger colonies will tend to be the correct clones. We have observed examples of either case when humanizing and bacterializing various loci. It is difficult to predict a priori which case will be evident for a given transformation, therefore it is often more practical to screen colonies and recording their size, and also ensuring that each size is adequately represented in the screen.
10. When picking colonies for the colony PCR screen, only a small quantity of cells is needed. Most likely as little as 1,000 cells will be sufficient to obtain a

PCR product. We have often chosen to collect slightly larger numbers of cells to visually confirm their suspension in water by turbidity. However, too many cells lead to incomplete lysis and inhibition of the colony PCR. With cell clumps larger than 1-2 mm the colony PCR will often fail. So ideally, the cells collected from the colony should form only a tiny speck, 0.5 mm or smaller in diameter. It is helpful to include the positive control PCR when screening, to identify samples which failed to produce a PCR product due to poor lysis. Lysis and PCR can be repeated for these samples if needed.

11. It is possible to adapt the protocol described here for the simultaneous replacement of multiple genes. The Mo Clo toolkit allows for cloning up to 4 different gRNA cassettes on the same CRISPR plasmid; for this, the gRNAs would be captured on pYTK050 as described here, but in the second Golden Gate reaction, instead of the ConL1 and ConRE plasmids, the first gRNA would be cloned with ConL1 and ConR2, the second with ConL2 and ConR3, the third with ConL3 and ConR4 and the fourth with ConL4 and ConRE (this process is explained in detail in Lee *et al.* 2015). All of these cassette plasmids would then be included in the final Golden Gate reaction to assemble the CRISPR plasmid. Then, during transformation of yeast, templates for each of the included gRNAs will need to be co-transformed. However, multiple replacements are even more dependent on efficient transformation, cleavage and repair than single replacements, and some additional work may be necessary to optimize these parameters in practice.

A.7. RECIPES

1. Zymolyase solution (50 ml). Weigh 9.11 g D-sorbitol. Dissolve in 50 ml distilled, deionized water to make 1 M sorbitol and autoclave. Weigh 0.25 g zymolyase and dissolve in sorbitol solution. Aliquot and store in -20 °C.
2. Lithium acetate, 100 mM (40 ml). Weigh 0.408 g lithium acetate dehydrate. Dissolve in 40 ml distilled, deionized water. Filter sterilize (0.2 μm filter) and store at room temperature.
3. LB medium (1 L). Weigh 25 g LB powder. For solid medium, add 15 g agar. Dissolve in distilled, deionized water for 1 L total volume. Autoclave and let it cool to 60-70 °C. Pour in petri plates so that the medium covers the visible area of the plate. Let plates cool and solidify at room temperature, store in 4 °C.
4. YPD (1 L). Weigh 50 g LB powder. For solid medium, add 20 g agar. Dissolve in distilled, deionized water for 1 L total volume. Autoclave and let it cool to 60-70 °C. Pour in petri plates so that the medium covers the visible area of the plate. Let plates cool and solidify at room temperature, store in 4 °C.
5. SD-Ura (1 L). Weigh 1.5 g yeast nitrogen base w/o amino acids, 5 g ammonium sulfate, 20 g dextrose, 2 g SC-Ura dropout powder. For solid medium, add 20 g agar. Dissolve in distilled, deionized water for 1 L total volume. Autoclave and let it cool to 60-70 °C. Pour in petri plates so that the medium covers the visible area of the plate. Let plates cool and solidify at room temperature, store in 4 °C.

A.8. FIGURES

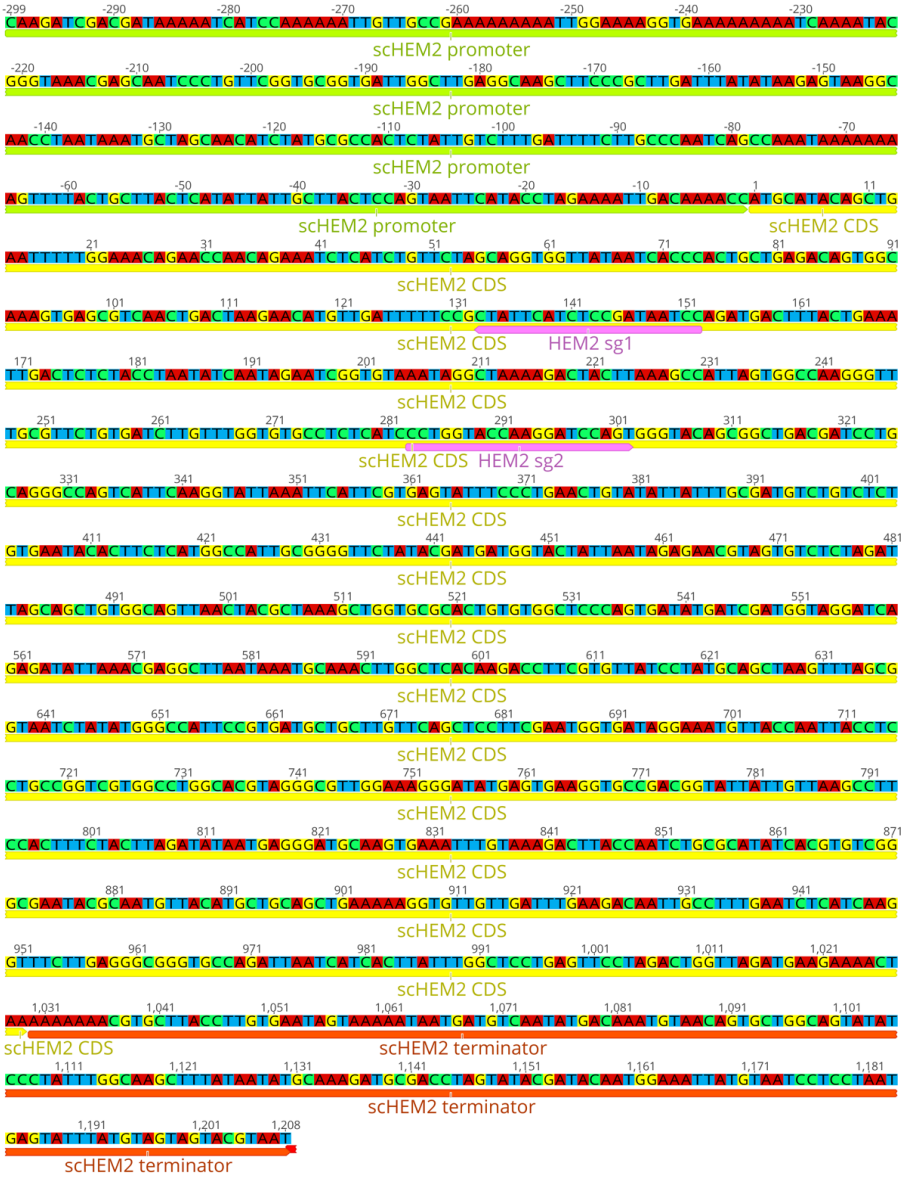


Figure A.8.1. Diagram of the native yeast HEM2 locus, showing positions of the example guide RNAs sg1 and sg2.

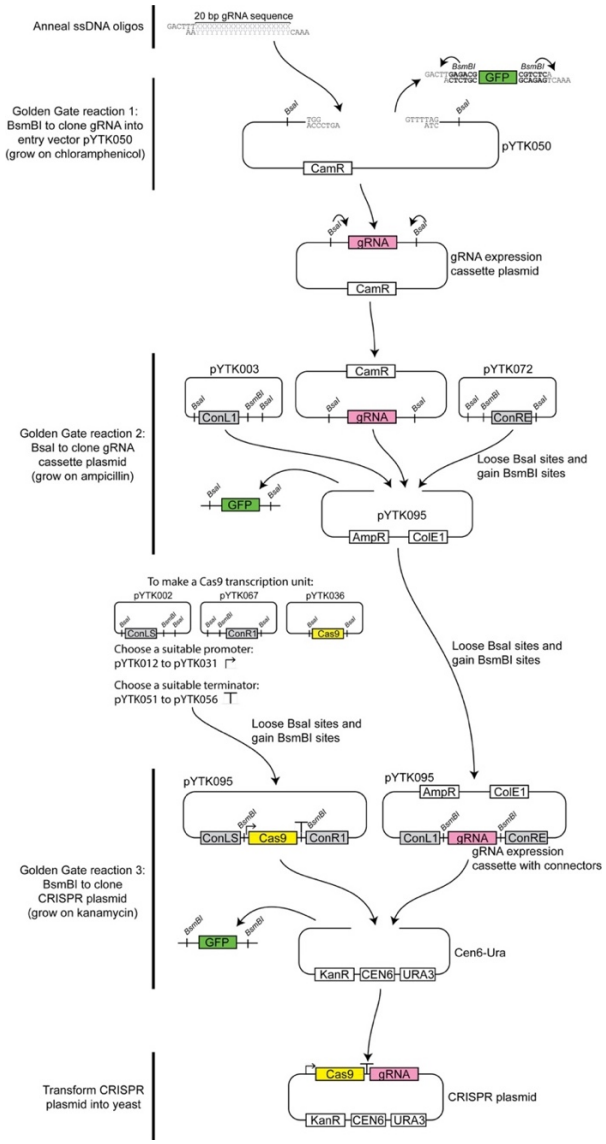


Figure A.8.2. Overview of the CRISPR plasmid construction process.

In the first step Xs and Ys represent the gRNA sequence selected, and BsmBI recognition site is indicated in bold.

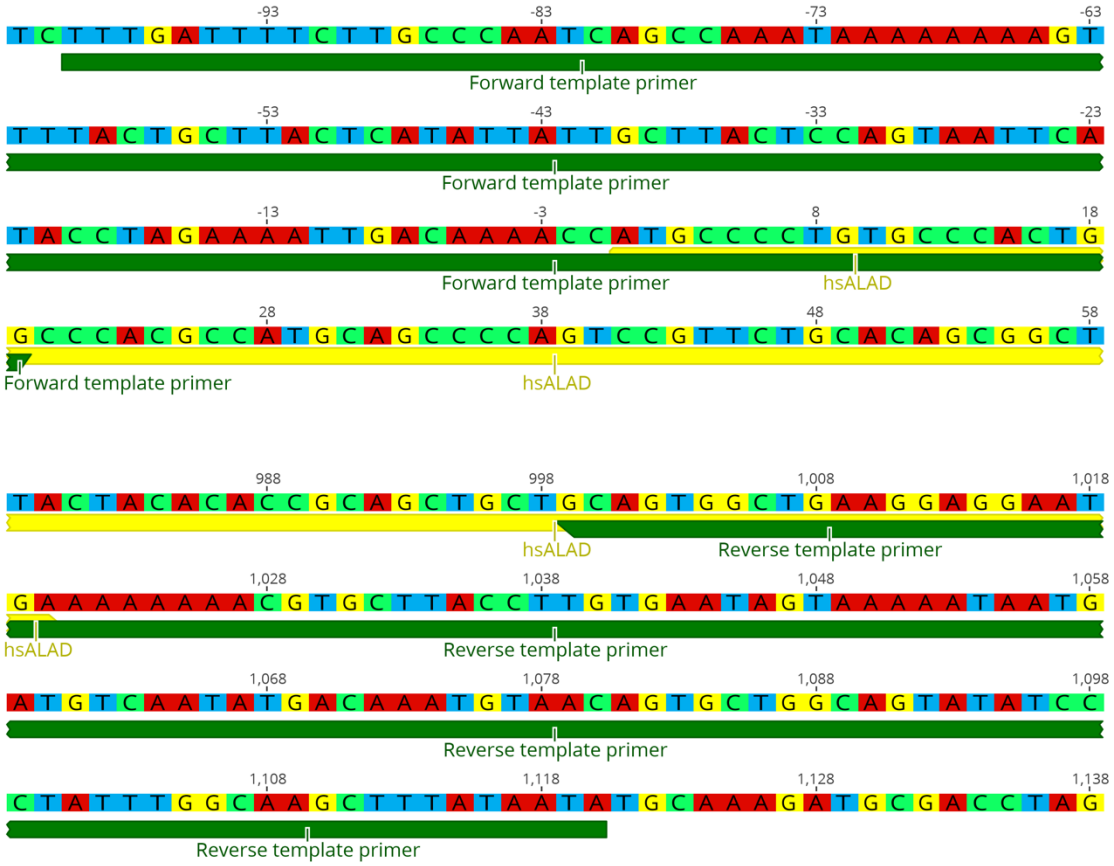


Figure A.8.3. Diagrams of example template primer designs for the replacement of HEM2 with hsALAD.

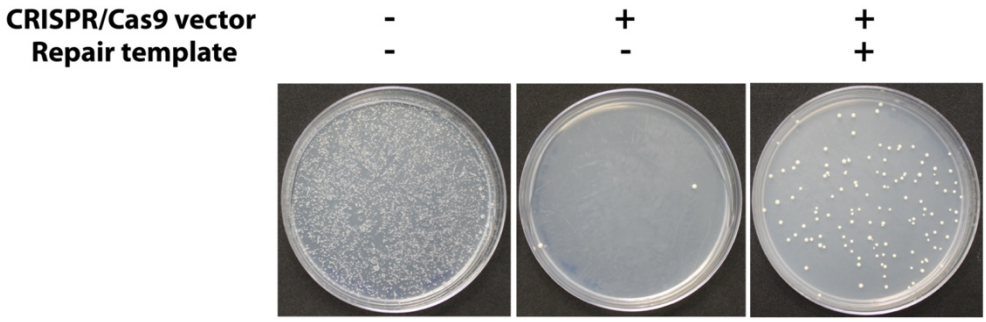


Figure A.8.4. Representative assay results.

Yeast cells are rescued from DSB lethality (Center plate) when an appropriate repair template is provided (Right plate). The left plate is a negative control of cells carrying a control plasmid with the same selectable marker (URA3) done to estimate the transformation efficiency of the yeast strains being used.

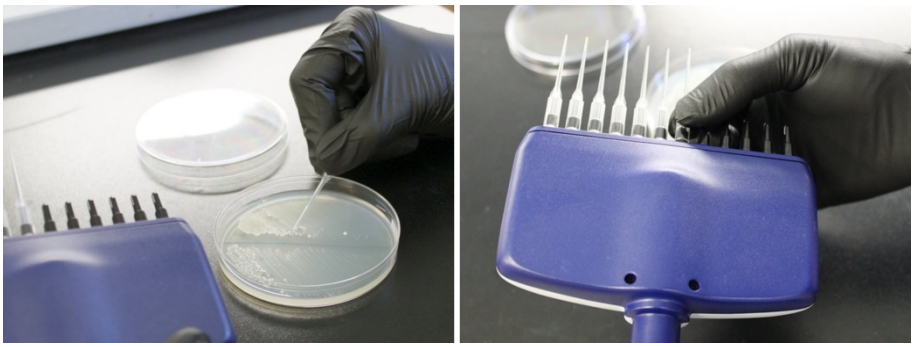


Figure A.8.5. Demonstration of colony picking technique with 12-channel pipette.

A.9. TABLES

Reagent	Amount
dsOligo	40 fmol
pYTK050	20 fmol
NEB T4 buffer 10x	1.0 µl
NEB T7 ligase	0.5 µl
NEB *Bsm*BI	0.5 µl
ddH_2O	To 10 µl

Table A.9.1. Golden Gate reaction for cloning into shuttle vector.

Reagent	Amount
gRNA on pYTK050	20 fmol
ConL1 (pYTK003)	20 fmol
ConRE (pYTK072)	20 fmol
AmpR-ColE1 (pYTK083)	20 fmol
NEB T4 buffer 10x	1.0 µl
NEB T7 ligase	0.5 µl
NEB *Bsa*I	0.5 µl
ddH_2O	To 10 µl

Table A.9.2. Golden Gate reaction for cloning gRNA cassette plasmid.

Reagent	Amount
gRNA cassette plasmid with connectors	20 fmol
Cen6-Ura cassette*	20 fmol
Cas9 plasmid (pYTK036)	20 fmol
NEB T4 buffer 10x	1.0 µl
NEB T7 ligase	0.5 µl
NEB *Bsm*BI	0.5 µl
ddH_2O	To 10 µl

*Cen6 Ura is constructed by assembling YTK plasmids (008, 047, 073, 074, 081, and 084).

Table A.9.3. Golden Gate reaction for cloning CRISPR plasmid.

Appendix B: Abundances of transcripts, proteins, and metabolites in the cell cycle of budding yeast reveal coordinate control of lipid metabolism[6]

One of the many strengths of the Marcotte lab is being able to use mass-spectrometry based proteomics to infer protein profiles and the complexes they assemble into in cells (Wan *et al.* 2015). During Michael Polymenis's sabbatical in the Marcotte lab, we extended this approach to infer changes in protein levels across the eukaryotic cell cycle. The eukaryotic cell cycle has been studied in detail for many years now but understanding the interplay between different biomolecules at the transcript, protein, and metabolite levels in the cell has been challenging for a number of reasons. Efforts have been predominantly limited by the ability to generate matched datasets to apply such multi-omics approaches. The problem is further complicated by the lack of synchronization and sampling bias across transcript and protein levels. Here, for the first time, we profiled the abundances of transcripts, proteins, and metabolites across the budding yeast cell cycle using multi-omics. The most interesting aspect of this project was generating experiment-matched datasets using cell size to efficiently sort and identify the cell cycle phases in yeast cells.

[6]This chapter was published as Blank, H.M., Papoulas, O., Maitra, N., Garge, R., Kennedy, B.K., Schilling, B., Marcotte, E.M., and Polymenis, M. (2020). Abundances of transcripts, proteins, and metabolites in the cell cycle of budding yeast reveal coordinate control of lipid metabolism. *MBoC* 31, 1069–1084. My contributions were planning and assisting with the proteomics experiments and helping with writing of the manuscript.

B.1. ABSTRACT

Establishing the pattern of abundance of molecules of interest during cell division has been a long-standing goal of cell cycle studies. Here, for the first time in any system, we present experiment-matched datasets of the levels of RNAs, proteins, metabolites, and lipids from un-arrested, growing, and synchronously dividing yeast cells. Overall, transcript and protein levels were correlated, but specific processes that appeared to change at the RNA level (e.g., ribosome biogenesis), did not do so at the protein level, and vice versa. We also found no significant changes in codon usage or the ribosome content during the cell cycle. We describe an unexpected mitotic peak in the abundance of ergosterol and thiamine biosynthesis enzymes. Although the levels of several metabolites changed in the cell cycle, by far the most significant changes were in the lipid repertoire, with phospholipids and triglycerides peaking strongly late in the cell cycle. Our findings provide an integrated view of the abundance of biomolecules in the eukaryotic cell cycle and point to a coordinate mitotic control of lipid metabolism.

B.2. INTRODUCTION

Exemplified by the discovery of cyclin proteins (Evans *et al.* 1983), identifying biomolecules whose abundance changes in the cell cycle has been a critical objective of cell cycle studies for decades. Recognizing such molecular landmarks in the cell cycle is a valuable, and often necessary, step for deciphering how and why cell cycle pathways are integrated.

Over the last twenty years, cell cycle-dependent changes in mRNA levels during the cell cycle of *S. cerevisiae* have been comprehensively defined not only from several arrest-and-release synchronization approaches (Cho *et al.* 1998; Spellman *et al.* 1998; de Lichtenberg *et al.* 2005; Pramila *et al.* 2006; Granovskaia *et al.* 2010), but also elutriation (Spellman *et al.* 1998; Blank *et al.* 2017). Unlike transcript profiling, cell cycle-dependent proteomic and metabolomic changes have been more limited and challenging to interpret due to different or poor synchronization, lack of matched transcriptomic datasets, and divergent results among the various studies. For example, there has only been one mass spectrometry-based proteomic analysis of the budding yeast cell cycle, sampling cultures at four time-points after they were released from arrest (Flory *et al.* 2006). Remarkably few proteins had altered levels during the time course of that experiment, and there was no correlation with the available transcriptomic datasets (Flory *et al.* 2006). Hence, at least in *S. cerevisiae*, it is not clear to what extent protein abundances are dynamic in the cell cycle, and how tightly they are linked to transcriptional changes, if at all.

The picture is not much clearer in other experimental systems. In fission yeast, two recent studies used highly similar arrest-and-release synchronization and protein labeling (stable isotope labeling by amino acids in the cell culture (Mann, 2006)) methods, followed by mass spectrometry, to probe cell cycle-dependent changes in the proteome. In one study only a single protein changed in abundance more than 2-fold (Carpy *et al.* 2014), while in the other report ~150 proteins did (Swaffer *et al.* 2016). Neither study had experiment-matched transcriptomic datasets. Previously, hundreds of

transcripts were reported to be periodic in the cell cycle of fission yeast (Rustici *et al.* 2004; Oliva *et al.* 2005).

In human cells, several reports sampled the proteome in the cell cycle with mass spectrometry, but there is little consensus among them (Dephoure *et al.* 2008; Olsen *et al.* 2010; Lane *et al.* 2013; Ly *et al.* 2015; Becher *et al.* 2018; Dai *et al.* 2018; Schillinger *et al.* 2018). The fraction of proteins identified as periodic ranged from ~5% (Ly *et al.* 2015), to >65% (Schillinger *et al.* 2018). Synchronization was mostly achieved by release from chemical arrest, but two studies also used elutriation (Ly *et al.* 2015; Dai *et al.* 2018). In the only report where an experiment-matched transcriptomic dataset was generated (Ly *et al.* 2015), the correlation with transcript abundance was positive (ρ=0.63, based on the Spearman rank correlation coefficient). Some of the differences among the above studies may arise from the use of different cell lines, such as: HeLa (Dephoure *et al.* 2008; Olsen *et al.* 2010; Lane *et al.* 2013; Becher *et al.* 2018); K562 (Dai *et al.* 2018); SW480 (Schillinger *et al.* 2018); or NB4 (Ly *et al.* 2015). However, even for the same cell line (HeLa), synchronization (release from thymidine block and nocodazole arrest), and point in the cell cycle (0.5 h after nocodazole arrest), the relative change in abundance of the 3,243 proteins identified in common between the two studies (Olsen *et al.* 2010; Becher *et al.* 2018) was poorly correlated (ρ=0.245, based on Spearman's rank correlation coefficient; see Materials and Methods).

In *S. cerevisiae*, metabolites have been measured in the cell cycle after arrest-and-release synchronization in minimal medium with ethanol as a carbon source, focusing on exogenous control of cell cycle progression and downstream effects on metabolism (Ewald *et al.* 2016). At the G1/S transition, it is generally thought that cyclin-dependent

kinase activity triggers lipolysis (Kurat *et al.* 2009) and mobilizes storage carbohydrates (Ewald *et al.* 2016; Zhao *et al.* 2016), to provide resources for cell division. In other systems, there is evidence of cell cycle-dependent changes on metabolite levels for the green alga Chlamydomonas reinhardtii (Jüppner *et al.* 2017), fly (Sanchez-Alvarez *et al.* 2015), and human HeLa cells (Scaglia *et al.* 2014; Atilla-Gokcumen *et al.* 2014; Ahn *et al.* 2017). Despite these advances, there has been no experiment-matched sampling of the transcriptome or proteome in any of these studies, making it difficult to integrate these datasets with gene expression, at the mRNA or protein levels.

Here, for the first time in any system, we generated comprehensive datasets for RNAs, proteins, metabolites, and lipids, from the same samples of S. cerevisiae cells progressing synchronously in the cell cycle. Importantly, these samples were from elutriated, un-arrested cells, maintaining as much as possible the normal coupling between cell growth and division. We found that while there is a broad correlation between the relative abundances of mRNAs and their corresponding proteins, cell cycle-dependent changes in transcriptional patterns are significantly dampened at the proteome level. The cellular lipid profile is highly cell cycle-regulated, with triglycerides and phospholipids peaking late in the cell cycle, together with protein levels of ergosterol biosynthetic enzymes, highlighting the importance of integrating multiple ‘omic’ datasets to identify cell cycle-dependent cellular processes.

B.3. MATERIALS AND METHODS

B.3.1. Strains and media

Unless noted otherwise, the cells were cultivated in the standard, rich, undefined medium YPD (1% $^{w}/_{v}$ yeast extract, 2% $^{w}/_{v}$ peptone, 2% $^{w}/_{v}$ dextrose), at 30 °C.

B.3.2. Elutriation

To collect enough cells for the downstream measurements of RNA, proteins, and metabolites, we followed the same strategy we described previously (Blank *et al.* 2017). Briefly, elutriated wild type, G1 cells (diploid BY4743 background) were allowed to progress in the cell cycle until they reached the desired cell size. At that point, they were quenched (with 100 µg/ml cycloheximide and 0.1% sodium azide) and frozen away, and later pooled with cells of similar size (Fig B.6.1A). Overall, we had to collect 101 individual samples, to generate the 24 pools shown in Fig B.6.1A.

For other elutriation experiments (e.g., see Fig.B.6.4 and 5), only an early G1 elutriated fraction was collected, from which samples were taken at regular intervals as the cells progressed in the cell cycle.

B.3.3. Cell size and DNA content measurements

The methods to measure DNA content and the cell size (birth or mean size) of asynchronous cultures and estimate the critical size of asynchronous cultures, have been described in detail previously (Guo *et al.* 2004; Truong *et al.* 2013; Soma *et al.* 2014; Maitra *et al.* 2019).

B.3.4. Proteomic samples

We used ~1E+09 cells from each of the 24 pools of the cell size series (see Fig B.6.1) to prepare extracts for LC-MS/MS. For each sample, the cells were resuspended in 0.75 ml of lysis solution (10 mM Tris pH 7.8, 10 mM NaCl). Glass beads were added to the top of liquid level, the samples were placed in a Mini Beadbeater (Biospec), and the cells broken by 'bead-beating' twice at the maximum speed for 90 s each time, placed on ice for 60 s between. The extract from each sample was collected by punching a hole with a 21-gauge syringe needle at the bottom of the tube. Lastly, the soluble material from the lysates were clarified by centrifugation at 14,000 *g* at 4 °C, for 10 m. Insoluble pellets were resuspended in 500 µl of lysis buffer and both clarified supernatants and pellets were stored at -80 °C until processing for mass spectrometry.

For mass spectral analysis, clarified extracts were thawed and protease inhibitors immediately added. 50 µl of each supernatant sample was mixed with 50 µl trifluoroethanol (TFE) and reduced with 5mM tris(2-carboxyethyl)phosphine (TCEP) at 56 °C for 45 m, cooled for 5 m at room temperature, and alkylated with 25 mM iodoacetamide in the dark, at room temperature for 30 m. Samples were diluted 10-fold with digestion buffer (50 mM Tris pH 8.0, 2 mM calcium chloride), digested with trypsin (added at 1:50 ratio) at 37 °C for 5 h. Digestion was stopped with 100 µl of 10% formic acid and sample volumes were reduced to 100-250 µl in a SpeedVac. Following filtration with an Amicon® Ultra-15 Centrifugal Filter Unit the peptides were desalted using C18 Spin Tips, according to the manufacturer's instructions. The volume of the samples was then reduced to 5-10 µl in a SpeedVac. Lastly, the samples were resuspended in 100 µl of

a 95% water, 5% acetonitrile, 0.1% formic acid solution, and subjected to LC-MS/MS analysis.

The insoluble pellets from the same extracts described above were processed based on a method reported previously (Lin *et al.* 2013). The pellets were resuspended in 50 µl of 2% $^{w}/_{v}$ sodium dodecyl sulfate (SDS), 50 mM ammonium carbonate and heated at 95 °C for 10 m. Following clarification each supernatant was transferred to a fresh tube, mixed with six volumes of cold acetone (-20 °C), and incubated at 4 °C for 4 h to form a precipitate. Precipitate was recovered by centrifugation at 13,000 g for 15 m, the supernatant carefully removed by aspiration, and the pellets washed twice with 0.4 ml of cold acetone. After each wash the samples were centrifuged at 14,000 g for 1 m and the supernatant carefully aspirated. Pellets were solubilized in 500 µl of 1% w/v sodium deoxycholate, 50 mM ammonium carbonate with two rounds of sonication (10 m each) in a water bath sonicator with 5 m on ice in between. 50 µl of each sample was reduced and alkylated with TCEP and iodocateamide as described above. Unreacted iodoacetamide was quenched with 12 mM dithiothreitol (DTT). The samples were brought to 80 µl with digestion buffer and digested with trypsin (added at 1:50 ratio) at 37 °C for 5 h. Digestion was stopped with 1% formic acid and samples were centrifuged at 14,000 g for 10 m to pellet the precipitated sodium deoxycholate. Peptides were desalted with C18 Spin Tips, and resuspended for LC-MS/MS as described above.

B.3.5. LC-MS/MS

Mass spectra were acquired on a Thermo Orbitrap Fusion. 5 µl (supernatant samples) or 2 µl (pellet samples) of peptides were separated using reverse phase chromatography on a Dionex Ultimate 3000 RSLCnano UHPLC system (Thermo

Scientific) with a C18 trap to Acclaim C18 PepMap RSLC column (Dionex; Thermo Scientific) configuration. Peptides were eluted using a 3-45% acetonitrile gradient over 70 min and directly injected into the mass spectrometer using nano-electrospray. Data-dependent tandem mass spectrometry was performed using a top speed HCD method with full precursor ion scans (MS1) collected at 120,000 m/z resolution and a cycle time of 3 sec. Monoisotopic precursor selection and charge-state screening were enabled, with ions of charge > + 1 selected with dynamic exclusion of 30 s for ions selected once within a 30 s window. Selected precursor ions underwent high-energy collision-induced dissociation (HCD) at 31% energy stepped +/-4%. All MS2 scans were centroid and done in rapid mode. Raw files were processed using Proteome Discoverer 2.2 and the label-free quantification workflow.

B.3.6. RNA samples and libraries

We used the same approach we had described previously (Blank *et al.* 2017), to collect cells from elutriated cultures of wild type (BY4743 strain background). For each of the 24 samples, from ~3E+07 cells total RNA was prepared with the hot phenol method. Briefly, the frozen pellets were re-suspended in 0.4 ml TES buffer (10 mM Tris pH = 7.5, 10mM EDTA, 0.5% SDS), in DEPC-treated water, and ~0.05 ml glass beads were added. Then, 0.4 ml of acid phenol:chloroform was added to each pellet, and the samples were incubated at 65 °C for 30 m, and vortexed briefly every 5 m during that time. The samples were centrifuged at 14,000 *g* for 5 m, and 0.3 ml of the top, aqueous layer were placed in a 2-ml screw-cap tube containing 1 ml cold ethanol with 40 µl of a 3M sodium acetate solution. The samples were incubated at 4 °C overnight and then centrifuged at 14,000 *g* for 20 m. The pellets were washed with 80% ethanol and

centrifuged at 14,000 *g* for 5 m. The pellets were air-dried and resuspended in 25 µl of DEPC-treated water. For the RNAseq libraries, we also used the same approach we had described (Blank *et al.* 2017), except that we did not select for polyA-tailed RNAs. Instead, from total RNA, we depleted rRNA, using the 'Ribo-Zero™ Magnetic Gold Kit (Yeast)', according to the manufacturer's instructions. All libraries were sequenced on an Illumina HiSeq4000, with multiplexing, at the Texas A&M AgriLife Genomics and Bioinformatics Facility. Raw sequencing data (fastq files) have been deposited (GEO: GSE135476).

The reads were aligned to the *S. cerevisiae* reference genome (version R64-1-1) using the *Rsubread* R language package (Liao *et al.* 2019). First, an index was built using the command: buildindex(basename = "R64", reference = "Saccharomyces_cerevisiae.R64-1-1.dna.toplevel.fa", gappedIndex=TRUE). Then, for each of the 24 libraries, the paired end reads were aligned with the command: align(index = 'R64', readfile1 = '….fastq.gz', readfile2 = '….fastq.gz', type = "rna"). For each library, we obtained >10 million uniquely mapped reads, and the output BAM files were then used in the featureCounts function of the *Rsubread* package, with the following command: featureCounts(files = "…subread.BAM", ispairedEnd = TRUE, requireBothEndsMapped = TRUE, annotext = "Saccharomyces_cerevisiae.R64-1-1.95.gtf", countChimericFragments = FALSE, isGTFAnnotationFile = TRUE).

For differential RNA levels between any two points in the cell cycle using the *DESeq2* R language package (Love *et al.* 2014). For this statistical analysis, the 24 cell size pools were grouped in 8 groups, for each of the approximately 5 fL increments in the

cell size series (see Fig B.6.1A). Additional analyses with ANOVA-based methods were performed as for the other biomolecules, and they are described below.

B.3.7. Metabolite samples and analysis

The untargeted, primary metabolite, biogenic amine, and complex lipid analyses were done at the NIH-funded West Coast Metabolomics Center at the University of California at Davis, according to their mass spectrometry protocols. Gas Chromatography–Time-of-Flight Mass Spectrometry (GC-TOF MS) was used for Primary metabolites. For biogenic amines, separation and detections was achieved by Hydrophilic Interaction Chromatography (HILIC), followed by Quadrupole time-of-flight (QTOF) MS/MS. Lastly, for complex lipids, Charged Surface Hybrid (CSH™) C18 separation was followed with QTOF MS/MS for lipids. Extract preparation was also done at the same facility, from 1E+07 cells in each sample, from the same ones used for proteomic and RNA profiling (Fig B.6.1). The cells were provided to the Metabolomics facility as frozen (at -80 °C) pellets. Detected species that could not be assigned to any compound were excluded from the analysis.

B.3.8. ANOVA-based computational approaches to identify differentially expressed biomolecules

For RNA samples, we used the TPM normalized values. For all other biomolecules, the input values we used were scaled-normalized for input values per sample. All the input and output datasets are shown in . To identify significant differences in the levels of biomolecules between any two points in the cell cycle we used the robust bootstrap ANOVA, via the *t1waybt* function, and the posthoc tests via the

mcppb20 function, of the *WRS2* R language package. For this statistical analysis, the 24 cell size pools were grouped in 8 groups, for each of the approximately 5 fL increments in the cell size series (see Fig B.6.1A).

B.3.9. SWATH-Mass spectrometry

The samples used to measure ribosomal protein abundances were from elutriated, diploid wild type BY4743 cells. Once the cells reached the desired cell size, they were quenched with 100 µg/ml cycloheximide and 0.1% sodium azide. Cells were harvested from three independently elutriated cultures (5E+07 cells in each sample). The cells were re-suspended in a buffer containing 20 mM Tris·Cl (pH 7.4), 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 µg/ml cycloheximide, 1% $^{v}/_{v}$ Triton X-100, and 25 U/ml Turbo DNase I, to a volume of 0.35 ml. Then, 0.2 ml of 0.5mm glass beads were added to each sample, and vortexed at maximum speed for 15 s, eight times, placing on ice for 15 s in between. The lysates were clarified by centrifuging at 5,000 rpm for 5 m, at 4 °C, and again for 5 m at 13,000 rpm at 4 °C. The supernatant was transferred to a 13×51 mm polycarbonate ultracentrifuge tube, underlaid with 0.90 ml of 1 M sucrose, and the ribosomes were pelleted by centrifugation in a TLA100.3 rotor (Beckman) at 100,000 rpm at 4 °C for 1 h. The protein pellets from three biological replicates for various time points during the cell cycle (40, 45, 50, 55, 60, 65, 70 and 75 fL) were then re-suspended in PBS, subjected to a Filter-Aided Sample Preparation (FASP) protocol tryptic digestion (Wiśniewski *et al.* 2009), desalted using C-18 Ziptips, and analyzed by data-independent acquisition (DIA)/SWATH-mass spectrometry, as described previously (Schilling *et al.* 2017).

Briefly, samples were analyzed by reverse-phase HPLC-ESI-MS/MS using an Eksigent Ultra Plus nano-LC 2D HPLC system (Dublin, CA) with a cHiPLC system (Eksigent) which was directly connected to a quadrupole time-of-flight (QqTOF) TripleTOF 6600 mass spectrometer (SCIEX, Concord, CAN) (Christensen *et al.* 2018). After injection, peptide mixtures were loaded onto a C18 pre-column chip (200 µm x 0.4 mm ChromXP C18-CL chip, 3 µm, 120 Å, SCIEX) and washed at 2 µl/min for 10 min with the loading solvent (H_2O/0.1% formic acid) for desalting. Subsequently, peptides were transferred to the 75 µm x 15 cm ChromXP C18-CL chip, 3 µm, 120 Å, (SCIEX), and eluted at a flow rate of 300 nL/min with a 3 h gradient using aqueous and acetonitrile solvent buffers.

For quantification, all peptide samples were analyzed by data-independent acquisition (Gillet *et al.* 2012), using 64 variable-width isolation windows (Schilling *et al.* 2017; Collins *et al.* 2017). The variable window width is adjusted according to the complexity of the typical MS1 ion current observed within a certain m/z range using a DIA 'variable window method' algorithm (more narrow windows were chosen in 'busy' m/z ranges, wide windows in m/z ranges with few eluting precursor ions). DIA acquisitions produce complex MS/MS spectra, which are a composite of all the analytes within each selected Q1 m/z window. The DIA cycle time of 3.2 s included a 250 ms precursor ion scan followed by 45 ms accumulation time for each of the 64 variable SWATH segments.

The DIA/SWATH data was processed with the Spectronaut™ software platform (Biognosys) for relative quantification comparing peptide peak areas among different time points during the cell cycle. For the DIA/SWATH MS2 data sets quantification was

based on XICs of 6-10 MS/MS fragment ions, typically y- and b-ions, matching to specific peptides present in the spectral libraries. Significantly changed proteins were accepted at a 5% FDR (q-value < 0.05).

B.3.10. Immunoblot analysis

For protein surveillance, protein extracts were made as described previously (Amberg *et al.* 2006), and run on 4-12% Tris-Glycine SDS-PAGE gels. To detect TAP-tagged proteins with the PAP reagent, we used immunoblots from extracts of the indicated strains as we described previously (Blank *et al.* 2017). Loading was evaluated with an anti-Pgk1p antibody.

B.3.11. Comparison of the relative protein abundances in (Becher *et al.* 2018) and (Olsen *et al.* 2010)

For the datasets generated in human, HeLa cells, 0.5 h after nocodazole arrest, the data were from Becher *et al.* (Becher *et al.* 2018) and Olsen *et al.* (Olsen *et al.* 2010). In the former study the authors reported the Log2-transformed ratios of the measured abundance over the median abundance of asynchronous cultures. For the (Olsen *et al.* 2010) proteins, the data were the isotopic ratios reported. In both cases, these values represented the corresponding protein abundances in that sample, among all the proteins identified in each sample in each study. To compare the rank order of the 3,243 proteins identified in common in the two studies, the Spearman's rank correlation rho (ρ) was estimated (ρ=0.245).

B.4. RESULTS

B.4.1. Samples for the multi-omic cell cycle analysis

To apply genome-wide methods for the identification of cell cycle-dependent changes in the abundance of molecules of interest, one must first obtain highly synchronous cell cultures. Preferably, synchronization must be achieved in a way that minimally perturbs cellular physiology and the coordination between cell growth and division (Aramayo and Polymenis 2017). When cells are chemically or genetically arrested in the cell cycle to induce synchrony, known arrest-related artifacts can bias the results (Ly *et al.* 2015; Aramayo and Polymenis 2017). An alternative synchronization method is elutriation, a physical process that fractionates an asynchronous cell population by cell size and sedimentation density properties of the cells, with minimal perturbation of cellular functions (Lindahl 1948; Creanor and Mitchison 1979; Banfalvi 2008). Hence, we used centrifugal elutriation to obtain our synchronous cell cultures (see Materials and Methods, and Fig B.6.1A). Elutriation separates cells primarily based on size, and size is used as a normalizing reference across different elutriation experiments. We isolated 101 different elutriated cultures, which were combined into 24 pools, based on the size at which they were harvested. Hence, we generated a cell size-series, spanning a range from 40 to 75 fL, sampled at approximately every 5 fL intervals. These 24 pools were processed as independent samples in all analytical downstream pipelines. For statistical analysis (e.g., with the bootstrap ANOVA), the 24 cell size pools were grouped in 8 groups, for each of the approximately 5 fL increments in the cell size series (see Fig B.6.1A). The same 24 distinct pools were aliquoted as needed (see Materials and Methods) to generate the input samples for measurements of RNA (with RNAseq),

proteins (with LC-MS/MS), and metabolites (GC-TOF MS for primary metabolites; HILIC-QTOF MS/MS for biogenic amines; and CSH-QTOF MS/MS for lipids).

To gauge the synchrony of our samples by microscopy, we used budding as a morphological landmark, which roughly coincides with the initiation of DNA replication in *S. cerevisiae*. The percentage of budded cells across the cell size series (Fig B.6.1B) rose steadily from ~0% in the smallest cells (at 40 fL), to >80% at the largest cell size (75 fL). The cell size at which half the cells were budded (a.k.a. 'critical size', a proxy for the commitment step START) in our cell size series was ~62 fL (Fig B.6.1B). This value is the same as the critical size these cells display in typical time-series experiments (Hoose *et al.* 2012). We also measured the DNA content of the cells with flow cytometry (Fig B.6.1). These results suggest that the synchrony is excellent for most of the cell cycle, except perhaps in the last 1-2 pools, when the cells have grown substantially in size and they are more heterogeneous. From the RNAseq data that we will describe later (Fig B.6.2), mRNAs that are known to increase in abundance at the G1/S transition (G1 cyclins; *CLN1,2*; peaking at 55 fL), or later in G2 phase (cyclin *CLB2*; peaking at 70 fL), peaked as expected in the cell size series (Fig B.6.1C). Hence, based on cytological (Fig. B.6.1B and S1) and molecular (cyclin mRNAs, Fig B.6.1C) markers of cell cycle progression, the synchrony of our samples was of high quality.

B.4.2. Overview of the datasets

One type of extract was analyzed for each class of the following biomolecules: RNA, primary metabolites, biogenic amines, and lipids (see Materials and Methods). For proteomic analysis, we used soluble protein extracts (designated as 'sol' in the datasets and material from the same extract that was recovered in an insoluble pellet (designated

as ‘pel’ in the datasets. The pellet was subsequently solubilized with detergents (see Materials and Methods) and analyzed in parallel to the soluble sample by liquid chromatography tandem mass spectrometry (LC-MS/MS). For label-free relative quantification of proteins, we used both spectral counts (designated as ‘psm’ in the datasets and peak areas (designated as ‘pa’ in the datasets. For RNAs, the signal we used for quantification was read counts, either raw or after normalization as Transcripts Per kilobase Million (TPM) (see Materials and Methods). For the metabolites, the signal was the peak heights from mass spectrometry (designated as ‘ph’ in the datasets).

For the quantification of proteins and metabolites, each dataset was first normalized for input. Hence, for proteins or metabolites, comparisons across the 24 samples were scaled based on the sum of the signals detected in each of the 24 samples. For RNA, we used TPM-normalized values and raw reads. For each dataset, we used a bootstrap-based ANOVA (see Materials and Methods; the output files named as ‘anova’ in the datasets. Also, for RNA, we used the *DESeq2* pipeline ((Love *et al.* 2014); see Materials and Methods; the output file designated as ‘deseq2’. Only biomolecules that changed ≥2-fold in our cell size series and had an adjusted p-value or FDR<0.05, were considered as significantly changing in the cell cycle.

For display purposes, in all the heatmaps and most plots, we show Log2-transformed expressed ratio values. These are the ratios of the levels that we measured for each biomolecule in each cell cycle point, reflecting the magnitude of the ratio of abundance relative to the average of that biomolecule across all the cell cycle points we sampled. This approach was originally used to describe microarray cell cycle experiments

in yeast (Spellman *et al.* 1998), and has been the standard in displaying and analyzing differential expression in the cell cycle.

B.4.3. RNAs in the cell cycle

The RNAseq data were analyzed (see Materials and Methods, Fig B.6.2,), to identify RNAs that change in abundance in the cell cycle. The number of identified RNAs varied, depending on the computational method. Based on the DESeq2 approach, ~40% of the transcripts (n=2,456) were significantly different between any two points in the cell cycle. The ANOVA-based approach identified 652 RNAs, whose levels changed significantly in the cell size series (Fig B.6.2). In addition to the expected clusters of RNAs associated with DNA replication (cluster 2) and mitotic cell cycle progression (cluster 4), there was a large cluster of transcripts enriched for processes related to ribosome biogenesis (cluster 1, Fig B.6.2;), peaking in the G1 phase. These transcripts also appeared periodic in past studies that relied on elutriation as a synchronization method to identify cell cycle-regulated RNAs (Spellman *et al.* 1998; Blank *et al.* 2017), but not in studies that used arrest-and-release methods (Spellman *et al.* 1998). An increase in the levels of transcripts involved in ribosome biogenesis before commitment to division has also been described in transcriptomic profiles of S. pombe (Oliva *et al.* 2005). Despite these changes at the transcript level, whether the ribosome content of the cell changes during the cell cycle is not known. We will describe results that do not support any cell cycle-dependent changes in assembled ribosomes (Figure 4).

Early in the cell cycle (cluster 1 & 3, Fig B.6.2), we noticed that there were some tRNAs whose levels were higher. Note that tRNAs were not examined in the two prior studies that queried the transcriptome of elutriated S. cerevisiae cells, because those

studies focused on polyA-tailed selected transcripts (Spellman *et al.* 1998; Blank *et al.* 2017). It has been argued that polyA selection biases the transcriptome quantification (Weinberg *et al.* 2016). Hence, in this study, we relied only on rRNA subtraction to prepare the RNAseq libraries (see Materials and Methods), which does not remove tRNAs and other non-coding RNAs. We also note that tRNAs are notoriously difficult to measure by RNAseq due to factors such as their high level of modification, sequence similarity between different tRNAs, and the difficulty to discriminate between cleaved and mature tRNAs. The tRNAs whose levels appeared to change in the cell cycle are shown in Fig B.6.2. These results are difficult to reconcile with the extreme stability of mature tRNAs (from 9 h to several days -exceeding the duration of multiple cell cycles, see (Hopper 2013)), unless these tRNAs are targets of quality control mechanisms (Hopper 2013). In any case, as we show later (Fig B.6.8) we found very little evidence to support a significant role for altered codon usage in the cell cycle.

Lastly, we also compared the identified cell cycle genes from the two elutriation datasets published previously (Spellman et al., 1998; Blank et al., 2017), and this study (Fig B.6.3). The overlap in the identified cell cycle transcripts appears low (e.g., the Spellman dataset overlaps ≈20% with either of the other two datasets). However, different methods in sample preparation (total RNA in this study vs. polyA-selected mRNAs in the other two) and in the computational analyses (ANOVA in this study vs. Fourier-transformation in the other two) were used, which could account at least in part for the differences.

B.4.4. Cell cycle-dependent changes in the proteome

From the soluble and insoluble extracts (see Materials and Methods), we identified 3,571 S. cerevisiae proteins, at one or more cell cycle points. From the soluble extracts 2,236 proteins were identified in all eight cell size groups in the cell cycle, while from the insoluble ones that number increased to 2,449 proteins. Although this represents a reasonably thorough sampling of the yeast proteome, we did not find some low abundance proteins (e.g., cyclins). This was not unexpected, since a recent, aggregate analysis of all available datasets of protein abundances in yeast (measured with tandem affinity purification (TAP), followed by immunoblot analysis-, mass spectrometry-, and GFP tag-based methods), placed proteins of the gene ontology process 'mitotic cell cycle regulation' as the least abundant group (Ho *et al.* 2018). The extent to which mRNA levels can explain protein levels is debated (Lu *et al.* 2007; Vogel and Marcotte 2012; Csárdi *et al.* 2015; Lahtvee *et al.* 2017). For most species, RNA levels explain between one to two-thirds of the variation in protein abundances (Vogel and Marcotte 2012). To examine the broad correlation between transcript and protein levels, we looked at the association of count data from our transcriptomic (reads) and proteomic (spectral counts) datasets (Fig B.6.4). Across all the points in our cell size series, the Spearman rank coefficients (ρ) for the transcriptome-proteome correlations ranged from 0.52 to 0.63 (Fig B.6.4). When we compared the relative abundance of the proteins in our datasets, against the consensus abundances calculated by (Ho et al., 2018), the correlation was higher ($\rho>0.75$ in every case, see Fig B.6.5). Hence, both in terms of the number of proteins we identified and their relative abundance, the quality of our proteomic datasets was high and consistent with past studies.

To identify proteins that changed in abundance in the cell cycle, we examined separately each of the four proteomic datasets: soluble and insoluble extracts, each quantified by spectral counts and by peak areas (see Materials and Methods). The overlap between the proteins in each dataset that appeared to change in abundance in the cell cycle was minimal (see Fig B.6.6). Based on ANOVA analysis, we identified 333 proteins whose levels changed significantly in the cell size series, in at least one of the four proteomic datasets (shown in the heatmap, in Fig B.6.3B). We will describe additional proteins whose levels change significantly in the cell cycle, but due to irregular patterns and missing values were not identified as such by the ANOVA-based method we used (see Fig B.6.5).

Our analysis provided numerous examples of physiologically relevant, cell cycle-dependent changes in protein abundance. Among these, were several whose levels are well known to be periodic at both the protein and RNA levels. These include proteins involved in DNA replication-related processes, such as both isoforms (Rnr1p and Rnr3p) of the large subunit of ribonucleoside-diphosphate reductase, peaking as cells enter S phase (Fig B.6.3A, bottom). However, other groups of proteins that we found to change in abundance in the cell cycle, were not so at the RNA level. For example, several enzymes of ergosterol biosynthesis (Erg1,11,3,5,7p) peaked late in the cell cycle (Fig B.6.3A, top). Of those, only the levels of the mRNA for Erg3p (C-5 sterol desaturase) changed in the cell cycle. The coordinate upregulation in the levels of enzymes involved in ergosterol biosynthesis is consistent with the mitotic increase in lipid levels that we will describe later (Fig B.6.6).

Despite the transcriptional upregulation in G1 of transcripts involved in ribosome biogenesis (see Fig B.6.2), we did not observe such broad changes at the proteomic level. In earlier reports, the synthesis of ribosomal components was not cell cycle-dependent (Shulman *et al.* 1973; Elliott *et al.* 1979; Warner 1999). To our knowledge, however, it is not known if the ribosome content in the cell, or the composition of ribosomal proteins in assembled ribosomes, changes in the cell cycle. Hence, we asked if the total amount of ribosomal proteins or their proportion in assembled ribosomes varies significantly in the cell cycle. To this end, we isolated assembled ribosomes through sucrose ultra-centrifugation from wild type cells (Fig B.6.4A; see Materials and Methods). Ribosomal protein abundance was measured with SWATH-mass spectrometry (see Materials and Methods). Note that for this experiment, extracts were not made from pools of different elutriated cultures, but from the same early G1 elutriated cells at different points as they progressed in the cell cycle (see Materials and Methods). The sum of all ribosomal protein abundances in assembled ribosomes did not change significantly in the cell cycle (Fig B.6.4B). Since our samples were prepared from the same number of cells, and ribosomal proteins accounted for ~60% of the intensities detected, one could reasonably assume that ribosome abundance does not change significantly in the cell cycle, but it is also possible that our isolation somehow has skewed the results. There were also no significant differences in the relative abundance of the individual ribosomal proteins in the cell cycle (Fig B.6.4C and S7). We note that the relative levels of each protein in the ribosomes of each sample are independent of any putative differences in the amounts of assembled ribosomes between samples. Hence, we conclude that gross ribosome composition is the same during the cell cycle. These results do not support, but also do

not unambiguously exclude, the possibility that individual, specialized ribosomes may be formed during the cell cycle. However, at least based on these population-averaged measurements, the composition of assembled ribosomes seems unaffected in the cell cycle.

Lastly, we interrogated our proteomic data for evidence of differences in codon usage during the cell cycle. It has been proposed that optimal codon usage is more prevalent in mRNAs expressed in the G1 phase of the cell cycle, contributing to the abundance of proteins that peak in G1 (Frenkel-Morgenstern *et al.* 2012). Altered tRNA abundances during stress conditions in *S. cerevisiae* may also regulate protein synthesis (Torrent *et al.* 2018). To avoid confounding effects from differential transcription of RNAs encoding the proteins that we identified to change in abundance in the cell cycle (Fig B.6.3B), we focused on the proteins whose corresponding mRNAs were not changing in the cell cycle (Fig B.6.2). Moreover, to minimize effects from regulated proteolysis, we excluded from the analysis proteins for which there is evidence for ubiquitylation and regulated proteolysis (Swaney *et al.* 2013). For the vast majority of codons in the remaining proteins, there were no significant changes between their actual and expected frequencies in the cell cycle, based on gene-specific codon usage (Tumu *et al.* 2012). Only four codons (AGC, UAU, AGG, AAC) were used with statistically significant differences in the cell cycle, but the magnitude of those differences was minimal nonetheless (Fig B.6.8). Overall, despite hints at the transcriptional level (Fig B.6.2) for upregulation of processes associated with protein synthesis in the G1 phase, at least from these population-based experiments, our data argue against any significant cell cycle-dependent changes in the ribosome content (Fig B.6.4B), composition (Fig

B.6.4C), or codon usage (Fig B.6.8), suggesting that at the proteome level those changes in RNA levels have been dampened extensively.

B.4.5. Thiamine biosynthesis and TDP-dependent enzymes in the cell cycle

To identify other proteins whose levels could change in the cell cycle but were not identified as such by the computational methods we used, we looked at proteins with the largest change in their levels, regardless of missing values or statistical cutoffs. Remarkably, a group of enzymes involved in thiamine biosynthesis peaked coordinately in abundance late in the cell cycle when the cells reached a cell size of ~65 fL (Fig B.6.5A). These enzymes participate in thiamine diphosphate (TDP) synthesis in the cytoplasm. Thi7p showed the smallest difference (slightly over 2-fold) in abundance during the cell cycle from our mass spectrometry experiments and could provide a good measure to validate our results. The *THI7* transcript was not identified in the core cell cycle regulated transcripts by Spellman et al (Spellman *et al.* 1998), which did not include in their analysis their own elutriation data. However, *THI7* was identified as cell cycle-regulated in all three elutriation datasets (this study, (Spellman *et al.* 1998), (Blank *et al.* 2017); and Fig B.6.3), suggesting that *THI7* expression is cell cycle-regulated at the RNA level, at least in experimental approaches that maintain the coordination of cell growth with cell division. To further test this prediction, early G1 cells carrying the *THI7-TAP* allele (the only available *THI* gene in the TAP-tagged strain collection encoding any of the proteins shown in Fig B.6.5A) were obtained by elutriation and the levels of the corresponding proteins were evaluated by immunoblotting at regular intervals, as the cultures progressed in the cell cycle (Fig B.6.5B). We confirmed by immunoblotting that the abundance of Thi7p was elevated late in the cell cycle (see Fig

B.6.5B; compared to the levels of the control protein Pgk1p). These results are consistent with the notion that there might be a coordinate, mitotic upregulation of thiamine biosynthesis enzymes.

Next, we asked if any TDP-dependent enzymes also change in abundance in the cell cycle and if strains lacking these proteins have cell cycle-related phenotypes. TDP is a cofactor for several enzymes, including transketolase (Tkl1,2p), α-ketoglutarate dehydrogenase (Kgd1p), E1 subunit of pyruvate dehydrogenase (Pda1p), pyruvate decarboxylase (Pdc1,5,6p), and phenylpyruvate decarboxylase (Aro10p). Only the levels of Tkl2p, Pdc5p, and Aro10p appeared to be elevated late in the cell cycle (Fig B.6.5C), at the same time as the levels of thiamine biosynthesis enzymes were also raised (Fig B.6.5A).

Cell size phenotypes are often used as a proxy for disrupted cell cycle progression with an increased cell size phenotype typically accompanying mitotic defects. Of all deletion strains lacking a protein that requires TDP as a cofactor, only the loss of Tkl2p increased cell size significantly (Fig B.6.5D). We found that both birth size and the mean size of *tkl2Δ* cells were larger (Fig B.6.5D). Note that the *tkl2Δ* deletion strain was not in the panels that were examined in genome-wide screens of cell size mutants (Jorgensen *et al.*, 2002; Zhang *et al.*, 2002). The mitotic upregulation in the levels of thiamine biosynthesis enzymes (Fig B.6.5A) and Tkl2p itself (Fig B.6.5C) are suggestive of possible mitotic roles for Tkl2p, which might depend on the available TDP pools in the cell. In the Discussion, we speculate on such putative roles, based on the published reports.

B.4.6. Cell cycle-dependent changes in metabolites and lipids

From the same elutriated pools we used to measure RNAs and proteins (see Fig B.6.1), we also measured metabolites and lipids. The assays were performed at the West Coast Metabolomics Center at UC Davis, an NIH RCMRC (Regional Comprehensive Metabolomics Resource Core). Each class of metabolites was measured with distinct mass spectrometry-based assays (see Materials and Methods). From these assays, thousands of compounds were detected, but most could not be assigned confidently to known metabolites, and they were not considered further. Instead, we focused on the 406 primary metabolites, biogenic amines, and complex lipids that were identified across the cell size series. As with our analysis of RNAs and proteins, we used ANOVA (Fig B.6.6) to identify compounds whose levels change in the cell cycle. Previous reports showed that storage carbohydrates are mobilized at the G1/S transition (Ewald et al., 2016; Zhao et al., 2016). In agreement with these studies, we also found that trehalose levels risein G1 to their highest levels when cell size reaches 50 fL but drop significantly at the G1/S transition (Fig B.6.6). The levels of other sugars (glucose, fructose, mannose, glucose-6-phosphate; see Fig B.6.6) were also higher in G1, consistent with an upregulation of glycolysis before DNA replication. By far, however, the class of metabolites that changed the most in abundance in the cell cycle was complex lipids, which peaked late in the cell cycle (Fig B.6.6). These included phospholipids (phosphatidyl-inositol (PI), -ethanolamine (PE), -serine (PS)) and triglycerides (Fig B.6.6). The higher triglyceride levels are also consistent with the elevated levels of neutral lipid droplets late in the G2/M phase, as reported previously (Blank et al., 2017). Overall, the coordinate increase in the levels of ergosterol biosynthesis enzymes we identified from the proteomic analysis (Fig B.6.3A) and the increase in lipids (Fig B.6.6), strongly suggest that lipid

metabolism is significantly upregulated late in the cell cycle. In the Discussion, we will expand on the significance of these results.

B.5. DISCUSSION

The sample-matched datasets for RNAs, proteins, metabolites, and lipids we generated from budding yeast cells progressing synchronously in the cell cycle provide a comprehensive view of these biomolecules in dividing cells. We discuss our approach to obtain the analytical samples, and our findings in the context of the relation between the transcriptome and the proteome and the integration of metabolite and lipid measurements with other 'omic' datasets.

A primary goal and unique aspect of this study are that we took a measure of the different biomolecules from the same, experiment-matched samples. Furthermore, to maintain the normal coupling between cell growth and cell division as much as possible, we used centrifugal elutriation to obtain synchronous cultures, instead of arrest-and-release methods. However, elutriation merely selects based on cell size a small fraction of cells from an asynchronous, larger population. Hence, compared to arrest-and-release methods, elutriation suffers from low yields of synchronous cells. To overcome this problem, we generated our samples by pooling from different elutriations cells of the same size. All the cells we collected from the elutriator were early G1 daughter cells, which were then allowed to progress in the cell cycle, before we harvested them once they reached a particular cell size. We then used cell size as a metric to arrange these

pools along a hypothetical cell cycle, generating a size-series (Fig B.6.1). This is different from starting with one elutriated sample, typically of small cells in early G1, and then taking aliquots at various times as they progress in the cell cycle, generating a time-series.

But practical necessities for large quantities of cells aside, how does a size-based series compare to a time-based one? We argue that it compares favorably, for the following reasons: First, no two, time series-based, elutriations are the same, and cell size always has to be used to account for experimental variation properly. A typical metric is the critical size for START, defined as the size at which half the cells are budded. We note that the critical size of our reconstructed cell cycle profile was ~62 fL (Fig B.6.1B), the same as the critical size of the same strain and medium in time-series experiments (Hoose *et al.* 2012). Second, based on DNA content analysis (Fig B.6.1), at least until the cells have grown substantially to ~65 fL, the synchrony appears very tight. Only at 70-75 fL, the samples appear more heterogeneous, with some cells going through the M/G1 transition, indicated by the rise of cells with G1 DNA content. But this is also a feature of time-based elutriations, or even arrest-and-release methods, where there is an inevitable loss of synchrony the further one moves along the cell cycle from the starting point. Third, every molecular marker we looked at (cyclin mRNAs, see Fig B.6.1C; proteins of DNA metabolism, see Fig B.6.3A) peaks as expected in our size-based series. Fourth, when comparing the cell cycle abundances of the same protein (Thi7p) from a size-based (Fig. C.6.5A) vs. a time-based series (Fig. C.6.5B), the profiles are nearly identical, even though the ploidy of the strains was not. For these reasons, our size-based approach recapitulates the expected properties of synchronous, elutriated samples.

In yeast, the latest meta-analyses from all available studies estimated that between 37% and 56% of the variance in protein abundance is explained by mRNA abundance (Ho *et al.*, 2018). These estimates are within the range of previous ones from multiple species (Vogel and Marcotte 2012). Based on the absolute quantification of protein and mRNA abundances (Lahtvee *et al.* 2017), the overall correlation between mRNA and protein abundances was also in that range (R^2=0.45, based on Pearson's correlation coefficient). The level of correlation between the transcriptome and the proteome we observed appears to be somewhat higher (ρ=0.52-0.63, based on Spearman's coefficient), probably because our experiments were done from synchronous cells, and because cell cycle transitions are associated with transcriptional waves (Spellman *et al.* 1998). A critical role for transcription in shaping the proteome takes place as cells transition in different environments, and during such transitions changes in protein levels were much more highly correlated with the changes in mRNA levels (R^2>0.9) (Lahtvee *et al.* 2017). Hence, the relatively high correlation we observed between the transcriptome and the proteome in the cell cycle is not surprising, and it is probably an underestimate, since some extremely unstable cell cycle regulators whose levels rise as a result of transcription (e.g., cyclins, see Fig B.6.1C), were absent from our proteomic datasets because of their low abundance.

Despite the correlation between the transcriptome and the proteome we discussed above, there were clear groups of transcripts and proteins whose abundance was incongruent. Ribosomal biosynthesis, reflected on the levels of individual ribosomal proteins or assembled ribosomes, was not periodic at the proteomic level (Fig. B.6.4 and S7), despite a large G1 transcriptional wave of RNAs involved in this process (Fig B.6.2).

We noted that a similar phenomenon was recently reported for the integrated stress response, a well-characterized transcriptional response in yeast involving ~900 transcripts (Gasch *et al.* 2000), which was not seen at all at the protein level (Ho *et al.* 2018). The observation that the ribosome content of the cell is constant in the cell cycle (Fig B.6.4) suggests that changes in translational efficiency of some mRNAs described previously (Blank *et al.* 2017) are likely due to transcript-specific mechanisms, rather than global changes in the steady-state ribosome content (Lodish 1974).

The mitotic peak in the levels of TDP biosynthesis enzymes was surprising (Fig B.6.5). The physiological significance of such a change in the levels of these enzymes is unclear. Through some uncharacterized roles, the TDP-dependent transketolase activity is necessary for meiotic progression in mouse oocytes (Kim *et al.* 2012). In bacteria, transketolase participates in chromosomal topology, and *E.coli* cells lacking transketolase are UV-sensitive (Hardy and Cozzarelli 2005). However, we found that yeast *tkl2Δ* cells are not sensitive to UV or other DNA-damaging agents (not shown). Overall, despite the intriguing observations that late in the cell cycle, levels of the TDP-dependent Tkl2p transketolase were higher (Fig B.6.5C) and loss of Tkl2p increased cell size (Fig B.6.5D), the molecular mechanism connecting these observations remains to be determined.

The coordinate upregulation of ergosterol biosynthetic enzymes late in the yeast cell cycle (Fig B.6.3), not evident at the RNA level (Fig B.6.2), was unexpected. To our knowledge, there is no prior report of such a response. It should be noted that the lack of cell cycle-dependent changes at the levels of mRNAs encoding the enzymes of ergosterol biosynthetis was seen not only in our dataset, but also in the other datasets aggregated in the Cyclebase 3.0 database for yeast and other organisms (Santos *et al.* 2015). Of the

enzymes we show in Fig B.6.3A, only *ERG3* had a rank score of 624, while all others were not periodic (scores >800) (Santos *et al.* 2015). Note that we also found *ERG3* mRNA levels to significantly change in the cell cycle.

The mitotic rise in the levels of sterol biosynthetic enzymes is significant in the context of our metabolite measurements, showing that lipid levels (especially phospholipids and triglycerides) increased at the same time (Fig B.6.6). Our observations are consistent with several other reports linking lipid metabolism with cell cycle progression and mitotic entry in yeast (Sd *et al.* 2012; McCusker and Kellogg 2012). Levels of triglycerides increase in wild-type cells synchronized in mitosis (Blank *et al.* 2017), storage of triglycerides in lipid droplets is thought to fuel mitotic exit (Yang *et al.* 2016), and lipid-exchange proteins integrate lipid signaling with cell-cycle progression (Huang *et al.* 2018a). Note that there have not been analytical measurements of distinct lipids in the cell cycle in yeast. The data we show here are not only consistent with, but also significantly expand the prior studies mentioned above. It is also important to stress that an increase in lipids late in the cell cycle cannot simply be due to a need for cell surface material. We had shown previously that increased lipogenesis does not increase cell size (Blank *et al.* 2017). Hence, the increase in the abundance of lipids likely reflects changes in the composition of membranes or other more specialized, cell cycle-dependent process, not necessarily a simplistic need for more cell surface building blocks.

One also needs to consider the dramatic changes in cellular morphology. Especially during mitosis, when the cell adopts the characteristic hourglass structure. The lipid content must accommodate dynamic changes in membrane curvature. For example, during cytokinesis, it is thought that lipids that confer negative curvature must be

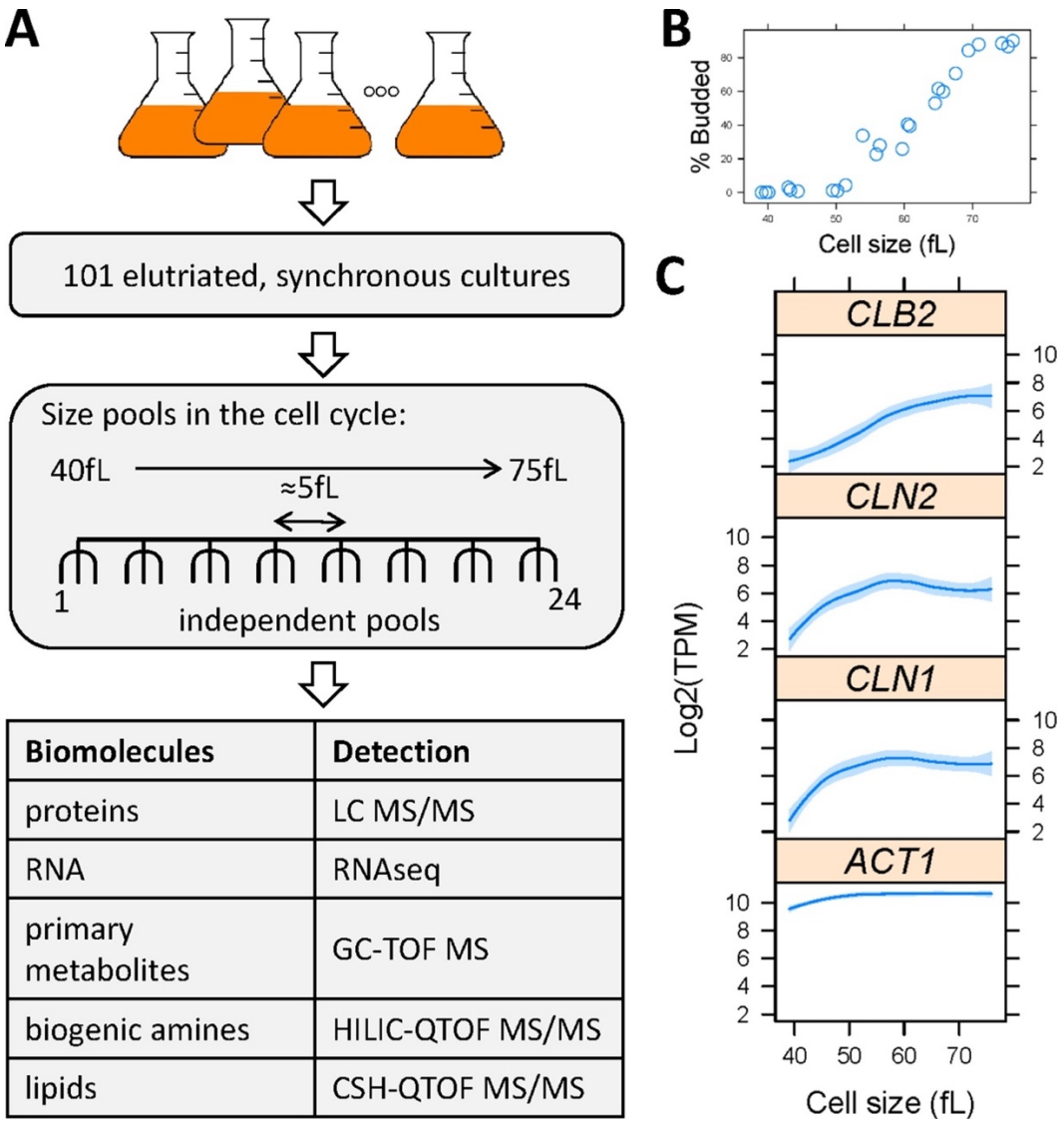
deposited on the outer leaflet of the bilayer (Furse and Shearman 2018). In yeast and human cells, inhibition of de novo fatty acid biosynthesis arrests cells in mitosis (Hasslacher *et al.* 1993; Schneiter *et al.* 1996; Al-Feel *et al.* 2003; Scaglia *et al.* 2014). In human cells, cholesterol synthesis may affect multiple points in the cell cycle. In an earlier report, inhibition of cholesterol synthesis arrested human cells in mitosis (Suárez *et al.* 2002), while in a later report the cells arrested in G1 (Singh *et al.* 2013). Cholesterol's role in mitosis appears to be complex, not only affecting the distribution of phospholipids in the plasma membrane but also governing the formation of a vesicular network at the midbody during cytokinesis (Kettle *et al.* 2015). Interestingly, ergosterol may have a cell cycle regulatory role in yeast, distinct from its bulk, structural role in membrane integrity (Dahl *et al.* 1987), but that role remains unclear (Gaber *et al.* 1989).

Lastly, our results argue for post-transcriptional mechanisms leading to mitotic upregulation of sterol biosynthesis. As to how the differential abundance of the ergosterol biosynthetic enzymes might come about, we note that all the enzymes we show in Fig B.6.3A, including Erg3p, have been shown to be ubiquitinylated (Peng *et al.* 2003; Swaney *et al.* 2013), raising the possibility of regulated proteolysis. Another possibility is translational control. In our previous analysis of translational control in the cell cycle by ribosome profiling, a confounding problem was the poor overlap of the different computational pipelines we used to analyze the data (Blank *et al.* 2017; Aramayo and Polymenis 2017). Proteomic data provide an independent metric that allows us to revisit the datasets of translationally controlled transcripts in the cell cycle. We found that for 38 transcripts (from Dataset 5 in (Blank *et al.* 2017)) identified by the *babel* pipeline (Olshen *et al.* 2013), and 45 transcripts (from Dataset 7 in (Blank *et al.* 2017)) identified

by the *anota* pipeline (Larsson *et al.* 2011), the corresponding protein levels are indeed significantly periodic in the cell cycle. These results validate the translational control of transcripts we had not followed up, but were identified by both computational approaches (e.g., *NQM1*, *MSC1*, with the protein levels peaking as predicted late in the cell cycle), and point to new ones, *ERG5* among them. Hence, multiple levels of control, such as transcription (for *ERG3*), translation (for *ERG5*), and protein degradation (possibly for all Erg proteins) may contribute to the mitotic upregulation of ergosterol biosynthesis. Hence, integrating our proteomic and ribosome profiling datasets increased the confidence for 5-fold more transcripts that are likely under translational control in the cell cycle, providing an example of the synergy among different studies and datasets.

Our data also underscore the value of having metabolite measurements along with other 'omic' datasets, to strengthen the efforts of identifying physiologically relevant cellular responses. In future work, employing targeted metabolic profiling and flux analysis in the cell cycle will increase our understanding of how the transcriptome and proteome shape dynamic changes in metabolism and how resources are allocated during cell division.

B.6. FIGURES



Biomolecules	Detection
proteins	LC MS/MS
RNA	RNAseq
primary metabolites	GC-TOF MS
biogenic amines	HILIC-QTOF MS/MS
lipids	CSH-QTOF MS/MS

Figure B.6.1. Overview of the experimental design to query cell cycle-dependent changes in the levels of RNAs, proteins, and metabolites.

(A) Generation of sample-matched, multiomic datasets from synchronous cultures of cells of different size, during the cell cycle. (B) Serving as a morphological marker of cell

cycle progression, the percentage of budded cells (y-axis) as a function of cell size (x-axis) is shown for each cell size pool. Cell size corresponds to the mean cell size of the population, and in this case it is the weighted average of all the mean cell sizes of all the elutriated samples that constituted each of the 24 pools. (C) The levels of mitotic (CLB2) or G1 (CLN1,2) cyclin mRNAs, which are known to be periodic in the cell cycle, are shown along with those of a non-periodic transcript (ACT1; encoding actin). Cell size is shown on the x-axis (in fL), while the Log2-transformed 'Transcripts Per Kilobase Million' (TPM) values for each transcript are shown on the y-axis. All 24 values, one for each pool, were plotted in these graphs. Loess curves and confidence bands indicating the standard errors on the curve at a 0.95 level were drawn using the default settings of the panel.smoother function of the latticeExtra R language package.

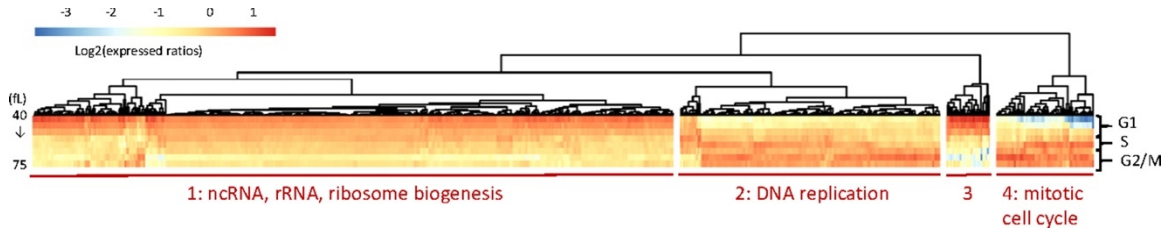


Figure B.6.2. Transcripts changing in abundance in the cell cycle.

Heatmap of the levels of 652 differentially expressed RNAs with significantly different levels ($p<0.05$; $Log2(FC)\geq1$) between any two points in the cell cycle, based on bootstrap ANOVA. The levels of each RNA were the average of each triplicate for the cell size indicated, which was then divided by the average value of the entire cell size series for that RNA. These 'expressed ratios' were then Log2-transformed. The Log2(expressed ratios) values were hierarchically clustered and displayed with the pheatmap R language package, using the default unsupervised algorithms of the package. The different rows of the heatmap correspond to the different cell sizes (40-75 fL, top to bottom, in 5fL intervals). The cell cycle phases approximately corresponding to these sizes are shown to the right of the heatmap. The gene ontology enrichment analysis for each cluster was done on the PANTHER platform.

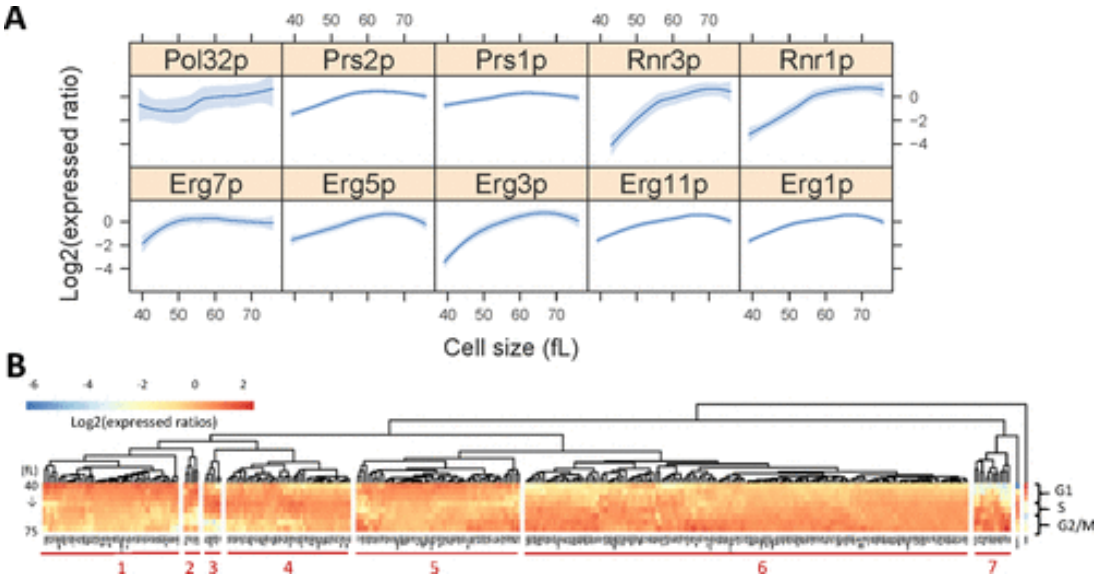
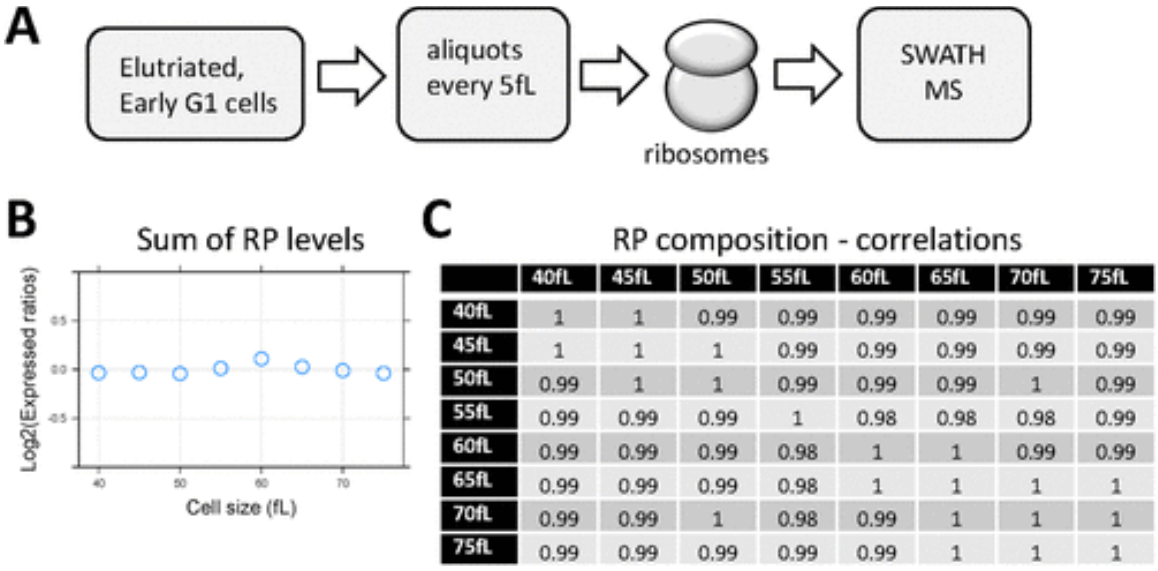


Figure B.6.3. Proteins with cell cycle-dependent abundance.

(A) Levels of selected proteins whose levels changed significantly ($p<0.05$; $Log2(FC)\geq1$) between any two points in the cell cycle, based on bootstrap ANOVA, in the cell cycle: Top, enzymes involved in ergosterol biosynthesis. Bottom, enzymes involved in DNA metabolism (Pol32p: DNA polymerase δ; Prs1,2p: PRPP synthase; Rnr1,3p: ribonucleoside-diphosphate reductase). The corresponding Log2(expressed ratios) values from all 24 data points are on the y-axis, and cell size values are on the x-axis. Loess curves and confidence bands indicating the standard errors on the curve at a 0.95 level were drawn using the default settings of the panel.smoother function of the latticeExtra R language package. (B) Heatmap displaying the relative abundance of the 333 proteins in one or more of the four 'anova' sets shown in Fig B.6.6. In cases where the same protein was in the intersection of more than one datasets, we chose for display the values from the dataset from which the changes in the protein abundance in the cell cycle was the

most significant (i.e., lowest p-value) and greater in magnitude (i.e., highest Log2(FC)). The heatmap was generated as in Fig B.6.2.



	40fL	45fL	50fL	55fL	60fL	65fL	70fL	75fL
40fL	1	1	0.99	0.99	0.99	0.99	0.99	0.99
45fL	1	1	1	0.99	0.99	0.99	0.99	0.99
50fL	0.99	1	1	0.99	0.99	0.99	1	0.99
55fL	0.99	0.99	0.99	1	0.98	0.98	0.98	0.99
60fL	0.99	0.99	0.99	0.98	1	1	0.99	0.99
65fL	0.99	0.99	0.99	0.98	1	1	1	1
70fL	0.99	0.99	1	0.98	0.99	1	1	1
75fL	0.99	0.99	0.99	0.99	0.99	1	1	1

Figure B.6.4. Ribosomal protein abundance in ribosomes does not change in the cell cycle.

(A) Elutriated, early G1 cells were cultured, and sampled at regular intervals in the cell cycle, in three biological replicates at each 5fL range, from 40 to 75 fL. Protein extracts from the same number of cells were then fractionated by sucrose ultra-centrifugation, to isolate ribosomes on mRNAs, which were then analyzed by SWATH-mass spectrometry (see materials and Methods). (B) The peak areas corresponding to each ribosomal protein (RP) detected were summed and averaged across the triplicate for each cell size interval. The Log2(expressed ratios) values for the 'Sum of RP levels' are shown on the y-axis, while cell size is on the x-axis. (C) Correlation matrix of the relative abundance of individual ribosomal proteins in assembled ribosomes on mRNAs. The Spearman correlation coefficients (ρ) shown in each case were calculated with the rcorr function of the Hmisc R language package. The cell cycle profiles for each ribosomal protein are shown in Fig B.6.7.

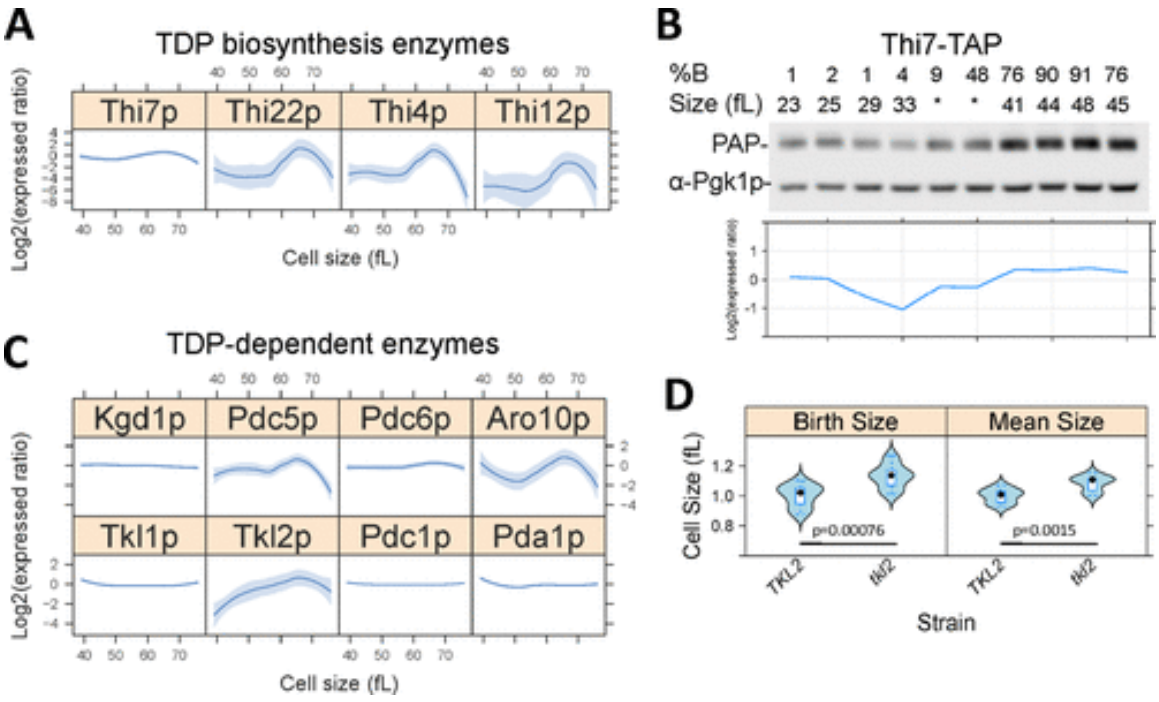


Figure B.6.5. Thiamine biosynthesis and TDP-dependent enzymes in the cell cycle.

(A) Abundances of the indicated proteins of thiamine biosynthesis from LC-MS/MS, across the cell size series (x-axis, in fL). The corresponding Log2(expressed ratios) values from all 24 data points are on the y-axis. Loess curves and confidence bands indicating the standard errors on the curve at a 0.95 level were drawn using the default settings of the panel.smoother function of the latticeExtra R language package. (B) The abundance of Thi7-TAP by immunoblotting from synchronous, elutriated cells, progressing in the cell cycle and sampled at regular intervals, as indicated (%B is the percentage of budded cells; fL is the cell size). Pgk1p levels are also shown from the same samples, to indicate loading. For the two samples indicated with asterisk (*) in the Thi7-TAP series, there were no size data due to instrument malfunction. At the bottom, the band intensities were quantified with ImageJ software, and the Log2-transformed

expressed ratios of Thi7-TAP are shown, after they were normalized against Pgk1p. (C) Abundances of the indicated TDP-dependent proteins, determined and displayed as in A. (D) The birth and mean size of tkl2 cells and experiment-matched wild type (TKL2) cultures from exponentially dividing cells in rich, undefined media (YPD). At least twelve independent cultures were measured in each case. Significant differences and the associated p values were indicated by the non-parametric Wilcoxon rank sum test, performed with the wilcox.test function of the R stats package.

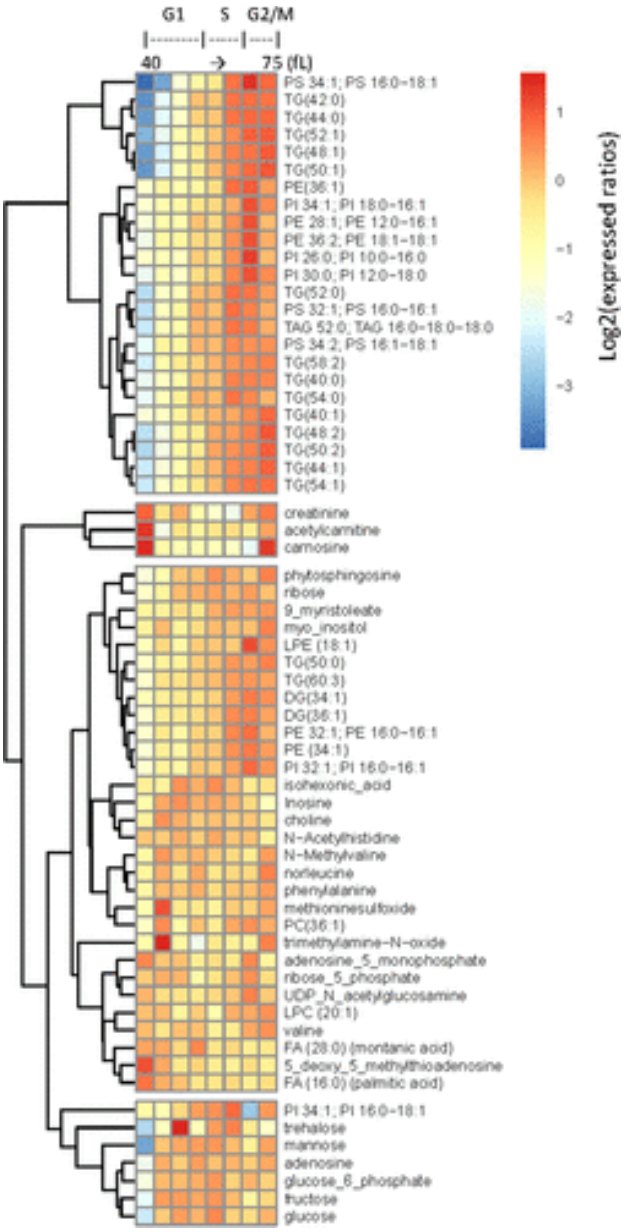


Figure B.6.6. Lipid levels change significantly in the cell cycle.

From 406 known metabolites identified from all classes (primary, biogenic amines, complex lipids), the levels of 64 with significantly different levels ($p<0.05$; $Log2(FC)\geq1$) between any two points in the cell cycle, based on bootstrap ANOVA, are shown in the

heatmap. The levels of each metabolite were the average of each triplicate for the cell size indicated, which was then divided by the average value of the entire cell size series for that metabolite. These ‘expressed ratios’ were then Log2-transformed. The Log2(expressed ratios) values were hierarchically clustered and displayed with the pheatmap R language package. The different columns of the heatmap correspond to the different cell sizes (40-75 fL, left to right, in 5fL intervals).

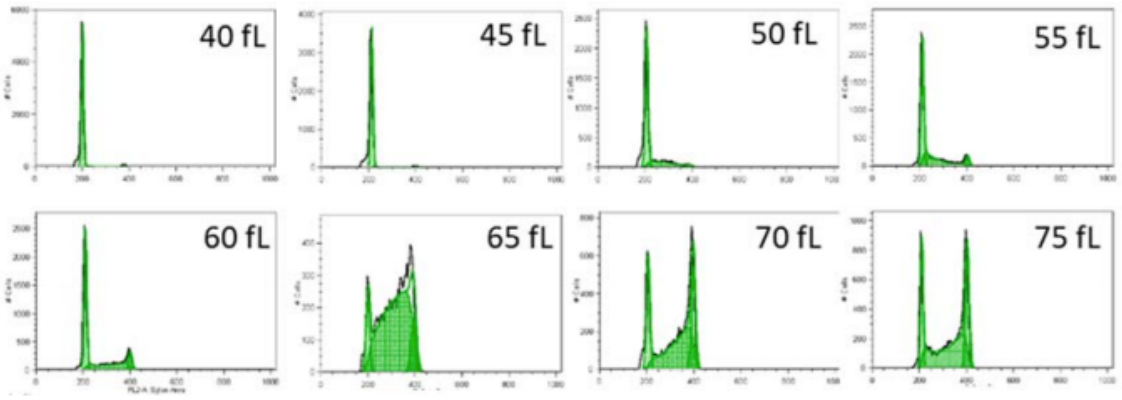


Figure B.6.S1. DNA content of samples spanning the cell size series from the elutriated samples.

The DNA was measured with flow cytometry, as described in the Materials and Methods. On the y-axis of each histogram is number of cells and on the x-axis the fluorescence per cell.

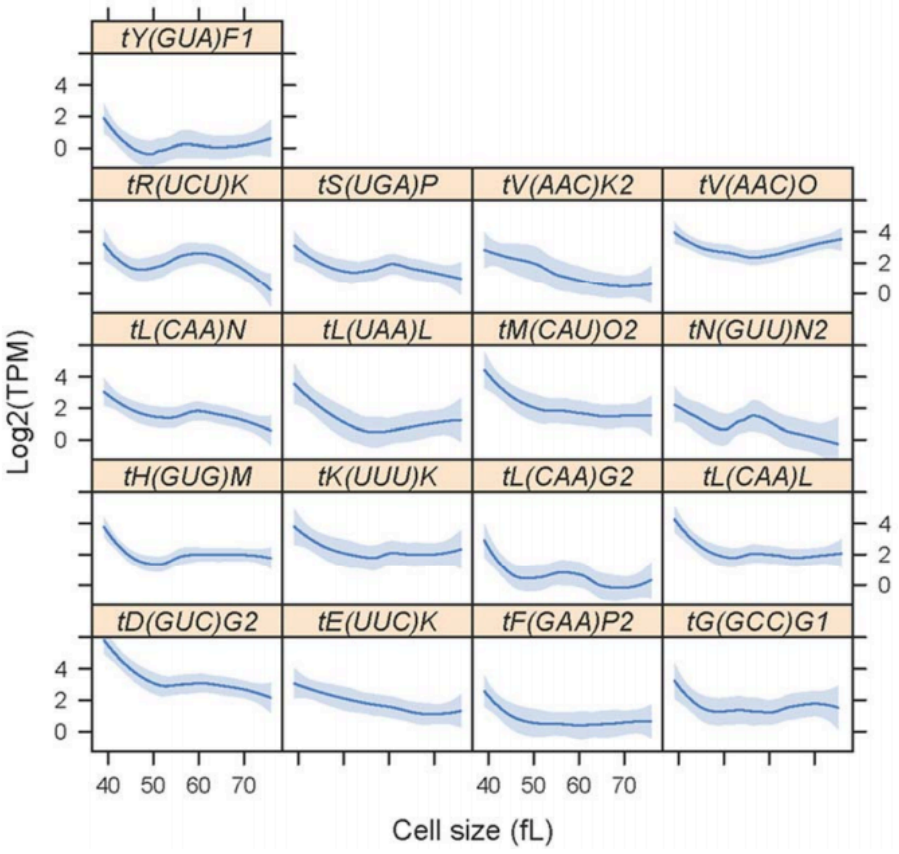


Figure B.6.S2. Levels of tRNAs, peaking early in the cell cycle.

The tRNAs were from clusters 1 and GURE3 in Fig B.6.2, with significantly different levels ($p<0.05$; $Log2(FC)\geq1$) between any two points in the cell cycle, based on bootstrap ANOVA. Sequences corresponding to the tRNAs shown peaked in abundance at cell sizes from 40 to 50 fL. Cell size is shown on the x-axis (in fL), while the Log2-transformed 'Transcripts Per Kilobase Million' (TPM) values for each tRNA from all 24 data points are shown on the y-axis. Loess curves and confidence bands indicating the standard errors on the curve at a 0.95 level were drawn using the default settings of the panel.smoother function of the latticeExtra R language package.

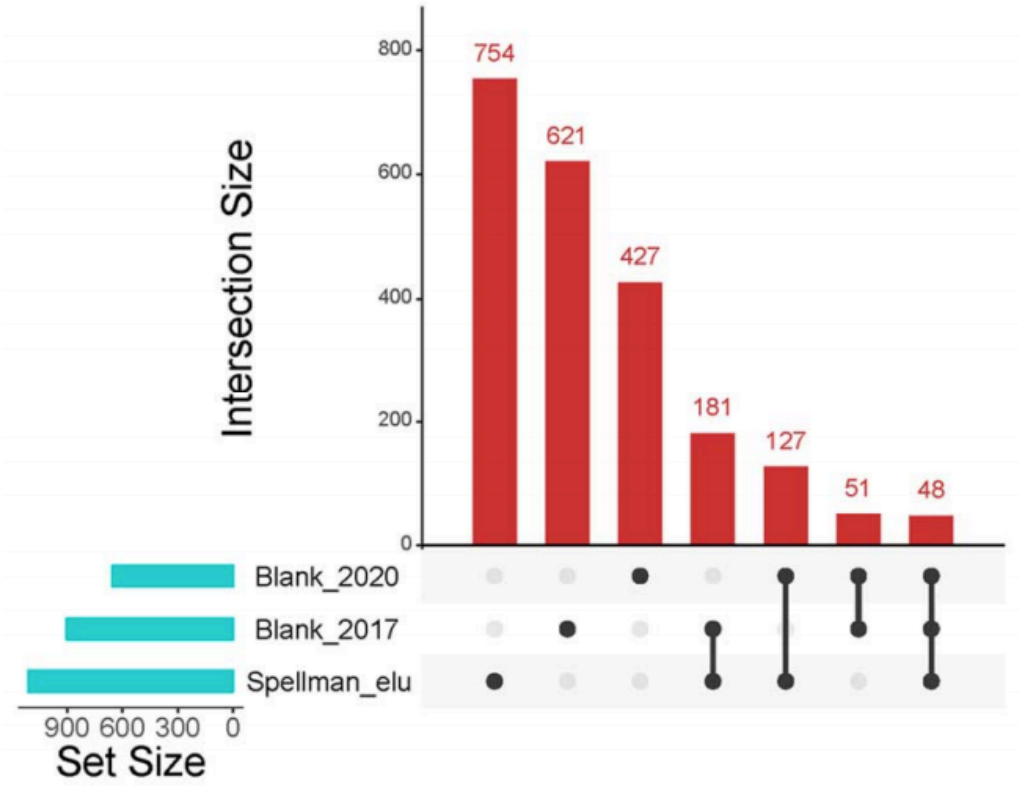


Fig B.6.S3. Overlap of transcripts whose levels change in the cell cycle, from studies that used elutriation to obtain synchronous samples.

Matrix layout for all intersections of the three sets. The graph was drawn with the UpSet R language package.

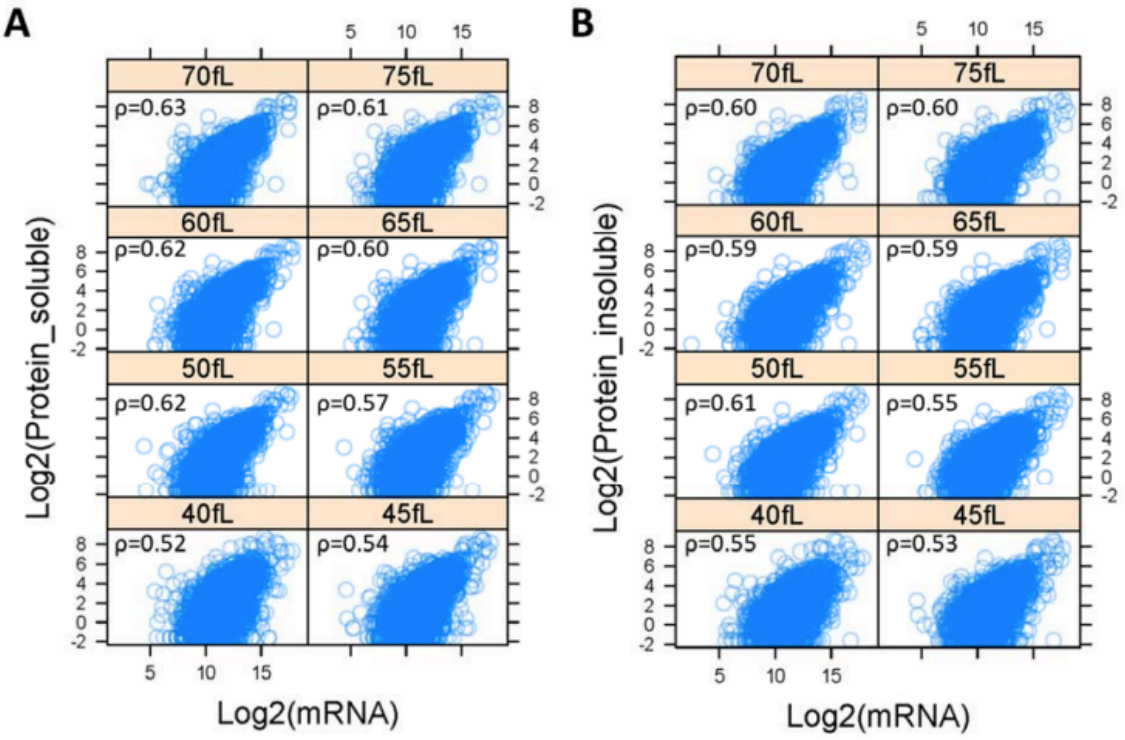


Figure B.6.S4. Transcriptome-proteome correlations.

(A) The spectral counts corresponding to the proteins identified in this study were averaged from the three biological replicates for each cell size pool we analyzed from the soluble fractions (from the 'sol_psm' dataset, and shown on the y-axis. On the x-axis are the RNA read counts from the corresponding loci. All values were Log2-transformed for display purposes. The Spearman correlation coefficients (ρ) shown in each case were calculated with the rcorr function of the Hmisc R language package. (B) Similar analysis as in A, except that the input dataset for the spectral counts (y-axis) was from the insoluble proteome fractions (from the 'pel_psm' dataset.

Dataset	40 fL	45 fL	50 fL	55 fL	60 fL	65 fL	70 fL	80 fL
soluble	0.81	0.79	0.79	0.81	0.80	0.80	0.80	0.80
insoluble	0.79	0.77	0.78	0.76	0.77	0.77	0.75	0.78

Figure B.6.S5. Relative protein abundances against literature consensus.

The protein abundances calculated in this study as in Fig B.6.4 for the soluble and insoluble fractions were compared to the consensus mean molecules per cell abundances from (Ho et al., 2018). The Spearman correlation coefficients (ρ) shown in each case were calculated with the rcorr function of the Hmisc R language package.

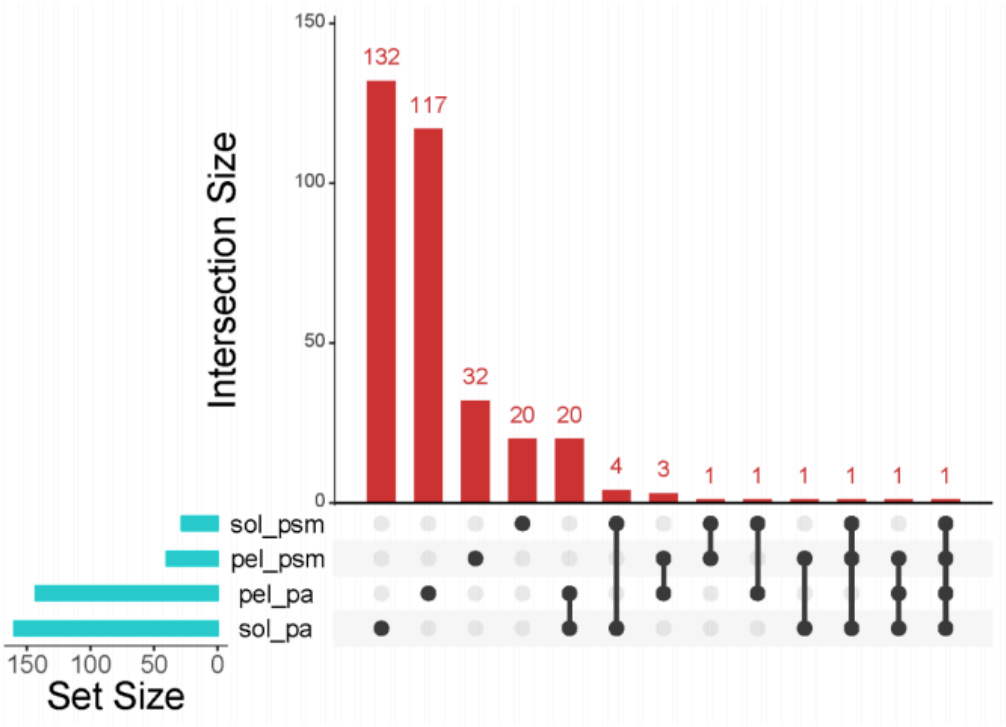


Figure B.6.S6. Overlap of protein datasets whose levels change in the cell cycle.

Matrix layout for all intersections of the four ANOVA-identified sets containing proteins with significantly different levels ($p<0.05$; $Log2(FC)\geq1$) between any two points in the cell cycle. The graph was drawn with the UpSet R language package.

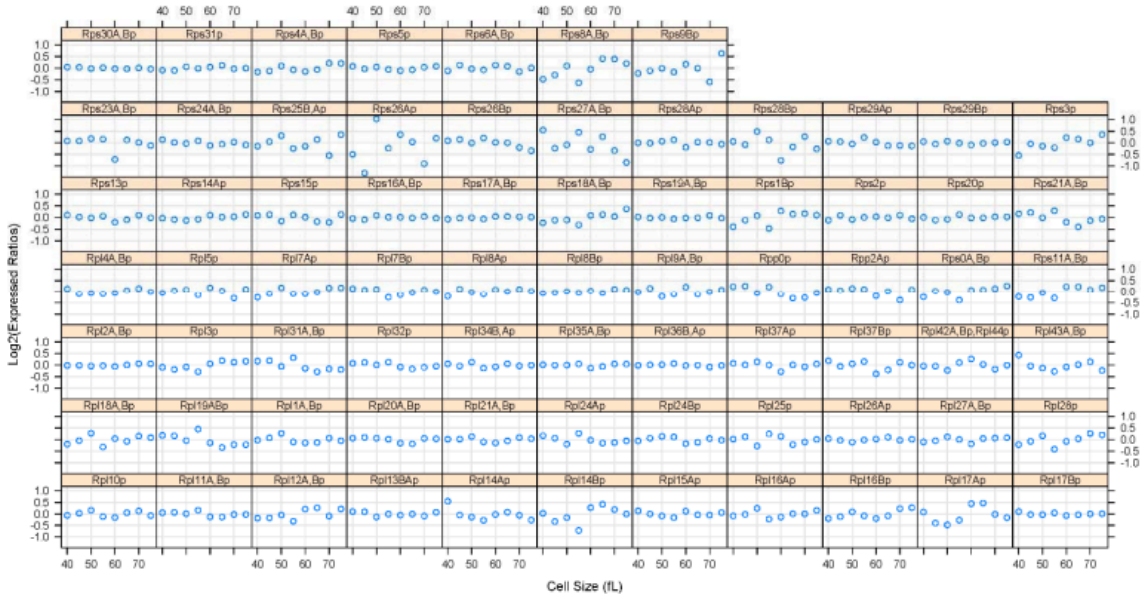


Figure B.6.S7. Ribosomal protein abundance in ribosomes is not periodic in the cell cycle.

The levels of each ribosomal protein (see Fig B.6.4) detected were normalized against the sum of all ribosomal proteins detected in that sample and displayed as Log2-transformed expressed ratios (y-axis), while cell size (in fL) is on the x-axis. In none of the few cases (e.g., Rps8,9,26,27,28p; Rpl14,17Ap) where the abundance of the ribosomal protein in question appeared to fluctuate somewhat in the cell cycle the changes were periodic (FDR>0.05), and these changes likely reflect experimental error in the quantification.

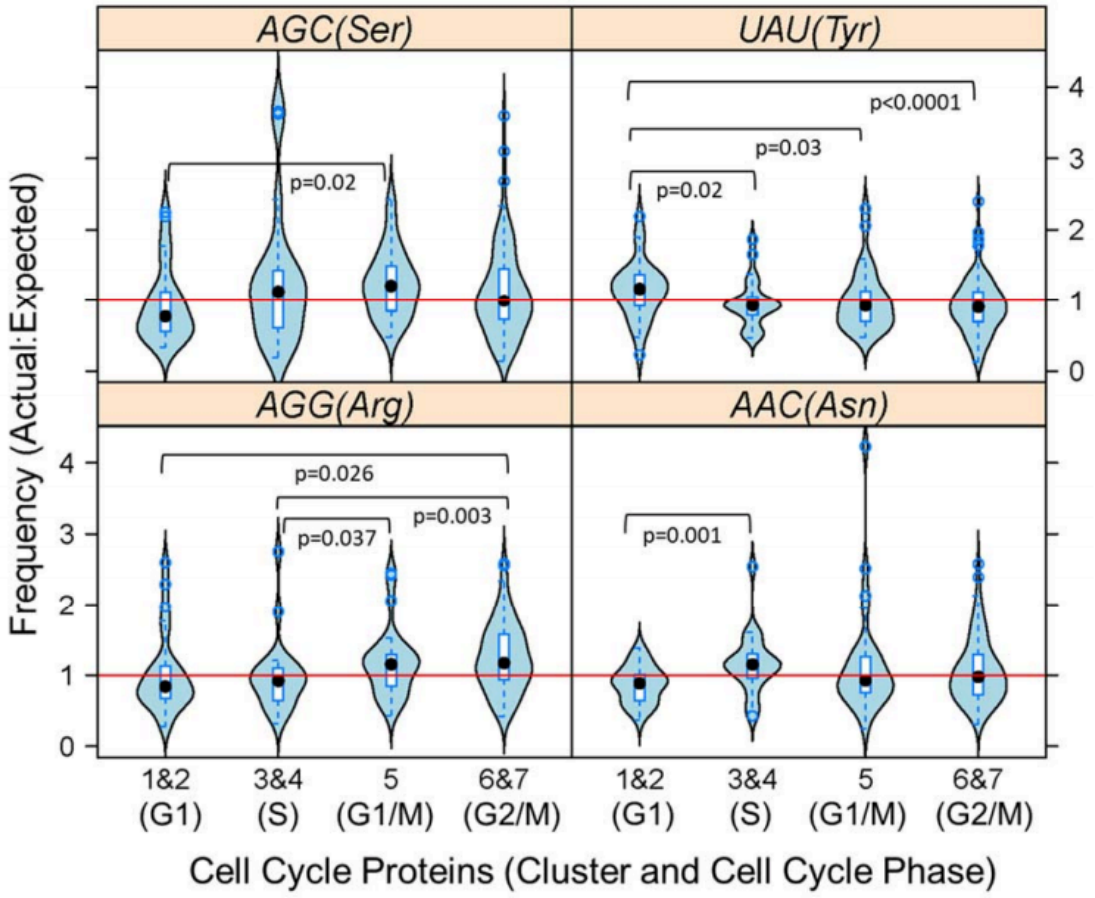


Figure B.6.S8. Little, if any, evidence for cell cycle-dependent changes in codon usage.

From the 333 cell cycle-regulated proteins shown in Fig B.6.3, we selected the ones who were not identified as ubiquitylated by (Swaney et al., 2013), and whose corresponding mRNA levels were not changing (from Fig B.6.2). These proteins were further grouped according to their cell cycle expression pattern (peaking in G1: in clusters 1&2 (n=29); peaking in S: in clusters 3&4 (n=24): peaking in G1/M: in cluster 5 (n=29); peaking in G2/M: in clusters 6&7 (n=90)). For each codon in each mRNA encoding each of these proteins, we obtained the ratio of the actual to expected usage, based on (Tumu et al.,

2012). These values are displayed as violin plots, for the four codons shown that there were statistically significant differences between the groups for each codon (based on bootstrapped ANOVA: $p<0.05$). For differences between groups in each codon, the p-values shown were obtained from posthoc statistical tests, using the mcppb20 function of the WRS2 R language package. The red horizontal lines indicate equal actual:expected codon usage in each case.

Appendix C: Spatiotemporal transcriptional dynamics of the cycling mouse oviduct[7]

This is chapter contains our investigations studying the cycling transcriptional dynamics underlying the mouse oviduct. Understanding temporal dynamics that control core processes in development has been of mutual interest both in the Marcotte and Wallingford labs well before I joined. This chapter aims to address a central gap in reproductive biology and development, particularly understanding the gene expression changes regulating the mammalian menstrual/estrous cycle. While this process is one of the most commonly occurring across mammals and central to birth, it remains quite understudied including in model organisms. This chapter aims to address that gap by spatiotemporally profiling gene expression changes occurring across the mouse oviduct. The most interesting finding from this work was that the significant remodeling of the oviduct was accompanied by lack of large-scale changes temporally across the estrous cycle. However, we found strong expression changes spatially across the anteroposterior axis particularly towards ciliary genes.

Led by Elle Roberson, I joined the project with Anna Battenhouse assisting with the analysis of the gene expression datasets both across the estrous cycle and between the sections of the mice oviduct.

[7]This chapter was published as Roberson, E.C., Battenhouse, A.M., Garge, R.K., Tran, N.K., Marcotte, E.M., and Wallingford, J.B. (2021). Spatiotemporal transcriptional dynamics of the cycling mouse oviduct. *BioRxiv*. This work was led by Elle C. Roberson. My contributions were assisting with RNAseq analysis and writing of the manuscript.

C.1. ABSTRACT

Female fertility in mammals requires the iterative remodeling of the entire adult female reproductive tract across the menstrual/estrous cycle. However, while transcriptome dynamics across the estrous cycle have been reported in human and bovine models, no global analysis of gene expression across the estrous cycle has yet been reported for the mouse, an important and highly tractable model animal. Here, we examined the cellular composition and global transcriptional dynamics of the mouse oviduct along the anteroposterior axis and across the estrous cycle. Though we observed robust patterns of differential gene expression along the anteroposterior axis of the oviduct, we found surprisingly few changes in gene expression across the estrous cycle, in marked contrast to other mammals. The absence of significant remodeling observed in the oviduct by histology was reflected in turn by relatively limited transcriptional dynamics across the estrous cycle. Thus, the mouse oviduct, unlike other mammals, is relatively stable in cellular composition and RNA abundances across the estrous cycle. We speculate that this is an evolutionarily derived state that may reflect the extremely rapid five-day mouse estrous cycle.

C.2. INTRODUCTION

The adult mammalian female reproductive organs – ovary, oviduct, uterus, cervix, and vagina – hold an interesting position in mammalian physiology because they constantly and repeatedly engage in a complex remodeling process more commonly associated with development. With every menstrual or estrous cycle, fluctuations in

ovarian steroid hormone secretion drive remodeling of these tissues, characterized by changes in epithelial morphology, fluid secretion, proliferation, apoptosis, and/or atrophy (Brenner and West 1975). However, while decades of research have revealed how these hormones fluctuate across the menstrual/estrous cycle in multiple mammals, we understand far less about the mechanisms by which these hormones drive cyclic tissue morphogenesis.

The oviducts are most well-known as the conduit between the ovary and uterus, but it is critical to note they also function as the site of fertilization and pre-implantation embryonic development (Stewart and Behringer 2012; Coy *et al.* 2012; Coy and Yanagimachi 2015). Interestingly, the oviducts display robust patterning along the anteroposterior axis, as the two major epithelial cell types – multiciliated cells (MCCs) and secretory cells – are differentially enriched (Agduhr 1927; Stewart and Behringer 2012; Barton *et al.* 2020). At the anterior oviduct close to the ovary, MCCs are highly enriched and responsible for capturing the ovulated oocyte(s), while at the posterior close to the uterus, MCCs are very sparse and act as sperm reservoirs (Talbot *et al.* 1999; Suarez 2016). The reduced proportion of MCCs in the oviduct posterior reflects the significant increase in the proportion of secretory cells, which are critical to pre-implantation development (Coy and Yanagimachi 2015). This anteroposterior pattern of the mouse oviduct epithelium is known to be established early in postnatal life and requires signaling from the underlying mesenchyme (Yamanouchi *et al.* 2010). How this pattern may change across the estrous cycle is not known.

In many species, the cellular basis of morphogenesis of oviduct MCCs across the estrous cycle was described *via* scanning electron microscopy studies decades ago

(Brenner 1969; Verhage *et al.* 1973; Shirley and Reeder 1996). In Rhesus monkeys, for example, anterior oviduct epithelial cells are cuboidal at the beginning of the cycle, but then grow in height and develop cilia at their apical surfaces as estrogen increases (Brenner 1969). As estrogen decreases and progesterone increases, the MCCs regress and de-ciliate (Brenner 1969). In mice, such cellular changes have not been thoroughly investigated, although it was shown that the wet weight of oviducts, as well as RNA and protein concentration, increases in the first half of the estrous cycle, and then subsequently decreases during the second half (Bronson and Hamilton 1970; Yamanouchi *et al.* 2010). More recently, there is evidence that ciliary beat frequency changes in response to estrogen and progesterone (Bylander *et al.* 2010; Shi *et al.* 2011). Perhaps surprisingly, while steroid hormone signaling in the oviduct is crucial for female fertility (Herrera *et al.* 2020; Winuthayanon *et al.*), we still lack a comprehensive view of how cycling steroids impact cellular morphogenesis in the oviduct, especially in the mouse.

A major hurdle to filling this knowledge gap is our very limited understanding of the transcriptional dynamics that underlie cyclical morphogenesis in this tissue. Transcriptomic approaches in human, swine, and bovine oviducts have investigated the impact of estrogen and progesterone, and some studies have also explored anteroposterior patterning in humans and bovine oviducts (Bauersachs *et al.* 2004; Hess *et al.* 2013; Cerny *et al.* 2015; Kim *et al.* 2018; Sowamber *et al.* 2020). However, the majority of these studies have focused on either the transcriptional response to fertilization and early embryo development or transcriptional signatures associated with progression of high grade serous ovarian carcinoma in the oviduct. The absence of a dynamic mouse oviduct

transcriptome across the normal estrous cycle represents a critical gap in our knowledge, especially given its potential utility as a model for understanding mammalian fertility.

Here, we examined both the cellular changes and global transcriptional dynamics along the length of the oviduct and across the estrous cycle. First, we provide quantitative data on the density of MCCs along the anteroposterior axis of the oviduct and we show that MCCs do not remodel across the estrous cycle in mice. In addition, we present 3' RNA-seq (Lohman *et al.* 2016) data for the anterior (approximate infundibulum) and posterior (approximate isthmus) oviduct at each stage of the estrous cycle. While transcript abundances vary strongly along the anteroposterior axis, our analyses suggest the estrous cycle has a surprisingly limited impact on transcription. Our data complement previous studies of transcriptional dynamics in the oviduct of other mammals and provide an important new resource for genetic studies of oviduct function in the mouse.

C.3. Materials And Methods

C.3.1. Mice.

6-8-week-old Swiss Webster female mice were obtained from Charles River and allowed to acclimate from travel for 1 week. Mice were housed in individually ventilated cages in a pathogen-free facility with continuous food and water, with a controlled light cycle (light from 7am-7pm). 7-9-week-old females were estrous cycle staged using standard vaginal cytology (Ajayi and Akhigbe 2020). Mice were humanely euthanized with extended CO_2 exposure followed by cervical dislocation, and female reproductive

tracts were dissected. All animal experiments were approved by the University of Texas at Austin Institutional Animal Care and Use Committee.

C.3.2. Tissue processing & immunofluorescence.

Dissected oviducts were carefully linearized (Fig. C.6.1A) by teasing apart the supercoils, gently affixing to a strip of index card to keep the tissue straight, then fixing in 4% paraformaldehyde for either 4-6hr at room temperature (RT) or overnight at 4°C. Linearized fixed oviducts were washed in PBS, and then incubated in 30% sucrose at 4°C overnight. Following a brief incubation in NEG-50 Frozen Section Medium (ThermoFisher), oviducts were embedded in NEG-50 using an ethanol/dry ice bath. 12μm frozen longitudinal sections were cut on a cryostat (Leica) and dried overnight at RT. Frozen sections were stored at -20°C.

Tissue sections were washed in PBS + 0.1% Tween20 (PBST) three times to remove NEG-50. Tissues were blocked for at least 30min at RT with 5% normal donkey serum + PBS (block buffer). Primary antibodies for cilia (mouse anti-acetylated tubulin, 1:1000 dilution, Sigma, cat# 6-11B-1) were diluted in block buffer and incubated on slides for 2hr at RT or overnight at 4°C. After washing three times with PBST, slides were incubated with Alexa-Fluor coupled secondary antibodies (goat anti-mouse 647, 1:1000 dilution, ThermoFisher) and DAPI (1:1000 dilution, ThermoFisher) for at least 30min at RT. After washing, slides were mounted with Prolong Gold (ThermoFisher), and allowed to cure at RT in the dark overnight. Oviduct sections were imaged on either a Zeiss LSM700 point scanning confocal microscope or a CSU-W1 spinning disk Nikon confocal.

C.3.3. Tissue sectioning and quantitation.

Two to four sections spaced at least 100µm apart were analyzed for cilia at four locations along the anteroposterior axis of the oviduct. Images from the the fimbria, anterior third (approximate infundibulum), middle third (approximate ampulla), and posterior third (approximate isthmus) were acquired and quantified. We quantified ciliary area as previously published (Roberson *et al.* 2020). Briefly, we used FIJI image processing to trace the lumen based on DAPI staining to approximate the luminal length. Similarly, the ciliated surface of the lumen was traced and measured based on acetylated tubulin staining. From these measurements, the % ciliated surface area was calculated for each image, and we plotted using GraphPad Prism 8.

C.3.4. RNA isolation and cDNA synthesis.

For 3' TagSeq (see below), dissected oviduct samples were collected in duplicate for each of the four stages across both anterior and posterior regions. For qPCR, samples were collected in triplicate for each estrous cycle stage across anterior, middle, and posterior thirds of the mouse oviduct. Following storage in RNA*Later* Storage Solution (Sigma, cat#: R0901) at -20°C, oviduct tissue was manually disrupted and the lysate was spun through a QIAshredder column (Qiagen, cat#: 79656) to fully homogenize. A Qiagen RNeasy mini kit (Qiagen, cat#: 74106) was used to harvest RNA for RNAseq, and the Qiagen RNeasy micro kit (Qiagen, cat#: 74004) was used to harvest RNA for qPCR. Total RNA was then either provided to the Genomic Sequencing and Analysis Facility at the University of Texas at Austin for 3' TagSeq, or cDNA was synthesized using the iScript Reverse Transcription SuperMix (BioRad, cat#: 1708841) for qPCR.

C.3.5. qPCR.

Most primers were designed from a database for mouse and human qPCR primers incorporated into the UCSC genome browser (Zeisel *et al.* 2013). Primers for *Foxj1* and *Msx2* were designed using Primer3Plus software. We confirmed specificity of primers by ensuring that they BLAST to no more than 1 site in the genome. There were four outliers to this BLAST assessment: all four ribosome primer sets blasted to more than 1 site in the genome, likely because there are numerous ribosomal pseudogenes scattered throughout the genome (Sisu *et al.* 2020). In addition, we only used primers whose melting curve displayed a single peak.

Primer sets were diluted from a stock (100μm in TE buffer) to 1μM in distilled deionized H_2O. 2μL of each primer set (in technical duplicates) was allowed to dry in the bottom of a well in a MicroAmp Fast Optical 96 Well Reaction Plate (Thermofisher, cat#: 43-469-06). 10μL of a master mix of cDNA (250pg/well), Applied Biosystems SYBR Select Master Mix (Thermofisher, cat#: 44-729-18), and distilled deionized water was added to each well, the plates were sealed with MicroAmp Optical Adhesive Film (Thermofisher, cat#: 43-119-71) and allowed to incubate at RT in the dark for at least 15min to rehydrate primer. Plates were run on a ViiA-7 Real-Time PCR system (ThermoFisher), and CT values were auto-determined by the ViiA-7 software. The standard $2^{-\Delta\Delta Ct}$ method was then used to determine fold change based on the geomean of three 'housekeeping' genes (*Hprt*, *Dolk*, and *Sra1*) (Schmittgen and Livak, 2008).

C.3.6. TagSeq.

Tissue samples were collected in duplicate for each of the four estrous stages across both anterior and posterior regions of the mice oviduct, accounting for 16 samples in total. Library preparation and sequencing for TagSeq (Meyer *et al.* 2011; Lohman *et al.* 2016), a form of 3' RNA sequencing, were performed by the Genomic Sequencing and Analysis Facility (GSAF) at The University of Texas at Austin. Total RNA was isolated from each sample by addition of Trizol (Thermo Fisher) and the sample was transferred to a Phasemaker tube (Thermo Fisher). Total RNA was extracted following the protocol supplied by the manufacturer and further cleaned up using a RNeasy MinElute Cleanup Kit (Qiagen). RNA integrity number (RIN) was measured using an Agilent Bioanalyzer and 100 ng of RNA was used for the TagSeq protocol. The fragmentation/RT mix was prepared and added to each RNA sample, then heated to 95°C for 2.5 minutes on a Thermocycler and immediately put on ice for 2 minutes. After cooling and addition of the template switching oligo and SmartScribe RT, the fragmented RNA reaction was incubated at 42°C for 1hr, 65°C for 15 min. Next an AmPure bead clean-up was completed for the cDNA before it was amplified to incorporate the Illumina sequencing primer site, followed by another cleanup. The remaining portions of the Illumina adapter (the i5 and i7 indices) were then added through an additional 4 cycles of PCR. Final libraries were quantified with PicoGreen then pooled equally for size selection using the Blue Pippin from 355-550 bp. Resulting libraries were sequenced using an Illumina HiSeq 2500 instrument (50-nt single reads).

C.3.7. Sequence data pre-processing.

Fastq datasets were initially processed to collapse duplicates based on TagSeq molecular barcodes (Matz). Sequencing data quality, both before and after TagSeq pre-processing, was evaluated using the FastQC tool (v0.11.9) (Andrews) and reports were aggregated with the MultiQC program (v1.0) (Ewels *et al.* 2016).

C.3.8. TagSeq data analysis.

Single-end pseudo-alignment was performed against the mouse transcriptome (GENCODE M23 transcript sequences (Frankish *et al.* 2019)) using kallisto (v0.45.0) (Bray *et al.* 2016) with options -l 200 -s 50 --single-overhang –bias. Downstream analysis of transcript abundance data was performed in R (v3.4.4) following protocols outlined in Bioconductor (Gentleman *et al.* 2004). The tximport package (v1.6.0) (Soneson *et al.* 2015) was first used to roll up transcript-level counts into gene-level counts provided to the DESeq2 package (v1.18.1) (Love *et al.* 2014). Before further analysis, count data matrices were filtered to remove genes with fewer than 1 read across all included samples. A number of models were analyzed to explore the oviduct location/estrous stage relationship: Posterior versus Anterior locations, first providing Early and Late data separately (n=8 each), then again providing all datasets (n=16); and Late versus Early, first providing Anterior and Posterior location data separately (n=8 each), then again providing all datasets (n=16). Differentially expressed gene results reported are those with maximum adjusted p-value 0.05 and log2 fold change greater than 1.0 or less than -1.0.

Gene ontology (GO) analysis was performed using topGO R package (v2.34.0) (Alexa, 2020) with GO database org.Mm.eg.db (v3.7.0). The topGO classic algorithm

and Fisher's exact test were used in count data mode. Input genes had maximum adjusted p-values of 0.10. Separate analyses were performed for up-regulated (log2 fold change 0.5 or higher) and down-regulated (log2 fold change -0.5 or lower) genes. The background gene universe consisted of observed genes used in the DESeq2 analysis.

Full DESeq2 and topGO results are provided in the supplementary zip file for GEO accession GSE164718.

C.4. Results And Discussion

C.4.1. Quantification of multiciliated cell density in the cycling mouse oviduct.

Oviducts in most mammals consist of two key epithelial cell types – secretory and multiciliated cells (MCCs) – and these are unevenly distributed across the anteroposterior axis. MCCs are enriched anteriorly and secretory cells, posteriorly (Agduhr 1927; Yamanouchi *et al.* 2010). While dynamic morphology changes across the estrous cycle have been described in the oviducts of many mammals (Novak and Everett 1928; Brenner 1969; Ferenczy *et al.* 1972; Verhage *et al.* 1973; Shirley and Reeder 1996), the issue has not been investigated thoroughly in the mouse. To stringently assess cellular remodeling of the anteroposterior axis of the mouse oviduct across the estrous cycle, we linearized the normally supercoiled oviduct by carefully teasing apart each coil along the length of the organ (Fig. C.6.1A), and generated longitudinal sections of linearized tissue. We then performed immunostaining for acetylated tubulin (Tub^{Ac}) to label MCC cilia, and DAPI to label nuclei (Fig. C.6.1B-E).

By examining fimbria, anterior (approximate infundibulum), middle (approximate ampulla), and posterior (approximate isthmus) regions of the oviduct, we clearly observed the anterior bias in MCC density (Fig. C.6.1B-E). We quantified this pattern by calculating the percentage of cilia that line the oviduct lumen, based on Tub^{Ac} staining as described previously (Roberson *et al.* 2020) (Fig. C.6.1F).

By performing parallel analyses on linearized oviducts from each stage of the estrous cycle, we found that this pattern displayed no temporal changes, with the anterior oviduct being significantly enriched for MCCs at all stages (Fig. C.6.1F). The absence of significant remodeling of MCCs across the estrous cycle in the mouse oviduct presents a marked contrast to other mammals (Brenner 1969; Verhage *et al.* 1973; Shirley and Reeder 1996). We hypothesize that this discrepancy may relate to the very short (4 to 5 day) mouse estrous cycle (Ajayi and Akhigbe 2020) and time required to eliminate and re-establish ciliated cells.

C.4.2. Determining spatiotemporal transcriptome dynamics in the mouse oviduct.

In addition to the cellular morphology changes, patterns of secretion, and bulk RNA and protein concentrations are known to change during the mouse estrous cycle (Bronson and Hamilton 1970). To ask if such changes have a transcriptional basis, we performed 3' TagSeq, a high-throughput RNAseq profiling method that captures and sequences the 3' ends of mRNA transcripts, enabling efficient estimation of relative gene expression (Meyer *et al.* 2011; Lohman *et al.* 2016). We performed two biological replicates of 3' TagSeq on isolated anterior (highly ciliated) and posterior (minimally ciliated) oviducts at each stage of the estrous cycle (Fig. C.6.2A). Principal component analysis (PCA) revealed that the majority of expression differences occur along the A/P

axis (72% of variance), while the estrous cycle accounts for much less variance (13%) (Fig. C.6.2B).

The lack of strong variation across the estrous cycle was striking, so we considered that our data might be underpowered to detect statistically significant trends across time (n=2 per estrous cycle stage). We therefore repeated our analyses, collapsing the estrous cycle stages into 'Early' (proestrus and estrus) and 'Late' (metestrus and diestrus). This approach improved differentiation between the major cycle phases; the majority of differentially expressed genes (DEGs) were still found along the A/P axis, not across the estrous cycle. For example, of the 1,758 DEGs enriched in the anterior oviduct, over 65% of them (1,158) were shared between Early and Late phases (Fig. C.6.2C). Similarly, of the 1,203 DEGs enriched in the posterior, 45% (542) were shared between Early and Late (Fig 2D).

A more granular view of these data using volcano plots illustrates this finding. In the anterior oviduct, we identified only 27 early and 14 late DEGs (Fig. C.6.2E). Similarly, modest enrichments were identified in the posterior, with 54 early and 40 late DEGs (Fig. C.6.2F). The magnitude of the expression changes was also modest, with the majority of DEG effect sizes (absolute value of the log2 fold change) of two or less (59% of estrous cycle DEGs, vs. 44% of A/P DEGs). This result contrasts starkly to the transcriptional dynamics in the oviducts of other organisms, including human, bovine, and swine, across the menstrual/estrous cycle (Bauersachs *et al.* 2004; Hess *et al.* 2013; Cerny *et al.* 2015; Kim *et al.* 2018). For example, in bovine anterior oviducts, 972 genes have been reported as enriched early and 597 enriched late (Cerny *et al.* 2015), while in

human anterior oviducts, 650 genes have been reported as enriched early and 683 enriched late (Hess *et al.* 2013).

In contrast to the muted transcriptional dynamics across the estrous cycle, we observed highly robust and spatially-restricted patterns of gene expression along the A/P axis of the mouse oviduct. We identified 1337 genes enriched in the anterior, and 770 in the posterior early in the estrous cycle (Fig. C.6.2G). Similarly, late in the estrous cycle, 1579 genes were enriched in the anterior and 975 in the posterior (Fig. C.6.2H). Thus, our data suggest that the mouse oviduct experiences relatively modest transcriptional changes as it progresses through the estrous cycle, but displays robust transcriptional differences along the A/P axis.

C.4.3. The mouse oviduct transcriptome is relatively stable across the estrous cycle.

While our finding of modest transcriptional variation across the estrous cycle in the mouse oviduct was consistent with the general lack of robust tissue remodeling, it was still surprising given the impact of cycling steroid hormones on oviduct function (Barton *et al.* 2020). To ask if the small number of DEGs we observed across time were enriched in particular classes, we performed Gene Ontology (GO) analysis (Alexa and Rahnenfuhrer 2020) on the early versus late DEGs. In the late stages, DEGs were subtly enriched for xenobiotic metabolism: xenobiotic metabolic process, cellular response to xenobiotic stimulus, and xenobiotic catabolic process (Supp. Fig. C.6.1B).

More interestingly, DEGs in the early stages were enriched for translation-related terms: peptide metabolic process, peptide biosynthetic process, translation, gene expression, and cellular amide metabolic process (Supp. Fig. C.6.1A). This enrichment for translation-related terms early in the estrous cycle is consistent with the known

concomitant increase in protein concentration (Bronson and Hamilton, 1971). In agreement with the GO analysis, manual annotation revealed an enrichment of large and small ribosome subunit genes expressed in the early estrous cycle compared to late (Fig. C.6.3A), with these genes displaying a 2 to 4-fold change in abundance (Fig. C.6.3B).

Further curation identified potentially interesting candidates based on their known or hypothesized function in female fertility. For example, Mki67, Ddit4, Atp1a1, Slc7a11, Mmp11, Mmp7, Spock1, and Wsb1 were each enriched at early stages, while Mamdc4, Gstp2, Gsta3, Susd2, and Ap1g2 were enriched late, either in the anterior, posterior, or both (Fig. C.6.3C). Among these, Mmp7 and Slc7a11 were especially interesting, because both were upregulated early and down regulated late, a pattern that is consistent with that observed in the cycling human oviduct (Hess *et al.* 2013). Mmp7 is part of the matrix-degrading enzyme family, and its down-regulation late in the human menstrual cycle is hypothesized to help maintain oviduct matrix integrity as the pre-implantation embryo travels through the organ (Hess *et al.* 2013). Slc7a11 is a glutamate/cysteine antiporter that is similarly regulated in the bovine oviduct (Cerny *et al.* 2015). The majority of these DEGs also displayed modest effect sizes, similar to the ribosomal genes (Fig. C.6.3D).

To independently verify the observed stability of the oviduct transcriptome across the estrous cycle, we performed qPCR for several of the DEGs. Consistent with the transcriptome-wide data, we did not observe statistically significant changes across the estrous cycle for most genes assayed (Fig. C.6.3E-K), but the trends in expression consistently reflected trends in our TagSeq data. The robust differential expression observed by qPCR between anterior and posterior (see next section) indicate that these

negative findings across the cycle do not reflect a lack of sensitivity in our assays. Rather, we conclude that the relatively modest remodeling observed in the oviduct by histology (Fig. C.6.1) is reflected by relatively muted transcriptional dynamics across the estrous cycle. While future work will be required, it seems reasonable to suggest that the stability of the mouse oviduct across the estrous cycle represents an evolutionarily derived state associated with the extremely rapid mouse estrous cycle, which at four to five days, is much faster than in other organisms (Brenner and West 1975).

C.4.4. The mouse oviduct displays robust transcriptional patterning along the anteroposterior axis.

To better understand the systems and processes differentially varying along the A/P axis, we again turned to GO Term analysis (Alexa and Rahnenfuhrer 2020). Consistent with the strong anterior enrichment of MCCs observed by histology (Fig. C.6.1), anterior-enriched DEGs were highly enriched for cilia-related processes (Supp. Fig. C.6.1C). In the anterior, complexes that function in MCCs, including intraflagellar transport, transition zone, CPLANE, and BBSome proteins (Garcia *et al.* 2018), as well as ciliary motility (Legendre *et al.* 2020) and MCC transcription factors (Lewis and Stracker 2020) were enriched (Fig. C.6.4A). To provide additional resolution to these patterns of expression for MCC-specific genes, we performed qPCR for several at each estrous cycle stage, using not just anterior and posterior regions but also the intervening middle region of the oviduct. We observed robust, statistically significant enrichment in the anterior region for a ciliary transcription factor (Foxj1), a transition zone complex gene (Tmem231), two ciliogenesis genes (Ift140 and Ift57), and one gene encoding a component of the ciliary beating machinery (Dnah9) (Fig. C.6.4F-I).

To investigate whether mouse anterior oviduct DEGs are similar to human multiciliated tissues, we compared our MCC-enriched dataset to the recently published human multiciliated tissue transcriptional signature (Patir *et al.* 2020). The analysis of four human multiciliated tissues, including ependymal, oviduct, trachea, and sperm, identified 248 genes expressed by all four tissues. Of those, we find 166 of them to be differentially expressed in the anterior of the mouse oviduct, suggesting that, transcriptionally, multiciliated genes of the mouse oviduct are highly similar to those of human multiciliated tissues.

Non-ciliated epithelial cells in the posterior oviduct consist predominantly of secretory cells (Agduhr 1927; Ghosh *et al.* 2017). Accordingly, our posterior-enriched gene set was strongly enriched for transport and secretion GO Terms (Supp. Fig. C.6.4), including solute carriers, transport ATPases, Ca^{2+} and K^{+} channels (Fig. C.6.4A). These GO terms are similar to previously published datasets from other mammals, including cows and humans, where vesicle-mediated transport, endocytosis, and exocytosis were among the main processes enriched in the posterior (Maillo *et al.* 2016; Gonella-Diaza *et al.* 2017; Rose *et al.* 2020). We confirmed this trend using qPCR for a transport ATPase (Atp1b1) and a solute carrier (Slc39a8); both were significantly enriched in the posterior (Fig. C.6.4K, L).

Interestingly, certain genes enriched in the mouse oviduct posterior are differentially expressed across the human menstrual cycle (Hess *et al.* 2013). For example, of the 33 solute carrier (Slc) genes that are enriched in the posterior of the mouse oviduct, eight of them are also differentially expressed across the human menstrual cycle, including Slc2a3, Slc22a23, Slc27a3, Slc39a14, Slc39a2, Slc39a8,

Slc4A7, Slc8a1 (Hess et al., 2013). These Slc genes are hypothesized to be important for secreting amino acids and other nutrients into the oviductal lumen to aid in preimplantation embryonic development. Further studies investigating SLC expression dynamics across both humans and mice may offer insights into the evolution of mammalian reproduction and divergence underlying these secretory cell types.

C.4.5. Patterned expression of known developmental signaling systems along the anteroposterior axis of the adult oviduct.

Our analysis of anteroposterior gene expression patterns also revealed several previously unreported trends. First, the genes up-regulated specifically in the posterior oviduct were strongly enriched for GO Terms related to embryonic development (Supp. Fig. C.6.1D). For example, all three non-canonical Wnt ligands, Wnt5a, Wnt7a, and Wnt11 were enriched in the posterior region of the oviduct (Fig. C.6.4B, N, O). Moreover, two antagonists of canonical Wnt signaling were also differentially expressed, but in a curious fashion: Dkk2 was enriched in the posterior, while Dkk3 was enriched in the anterior (Fig. C.6.4B). Components of the FGF signaling pathway were also differentially expressed: Fgf1, Fgf16, and Fgf2 were enriched in the posterior, while only the FGF receptor, Fgfr3, was enriched in the anterior (Fig. C.6.4C). Finally, multiple transcription factors were differentially expressed, including Msx2 (Fig. C.6.4M) and several Hox genes, including Hoxa9 and Hoxa10 (Fig. C.6.4D), both of which are involved in female reproductive tract development (Du and Taylor 2016). Other differentially expressed Hox genes include Hoxb2, Hoxb7, Hoxb8, and Hoxc10 (Fig. C.6.4D). We confirmed the differential expression of many of the developmental regulators using qPCR (Fig. C.6.4M-O)

Not all signaling pathways displayed patterned expression, and an excellent example is the Planar Cell Polarity (PCP) pathway (Fig. C.6.4E). The PCP system is essential for normal polarized ciliary beating in the oviduct, so most PCP genes are expressed in the oviduct (Shi *et al.* 2014; Koyama *et al.* 2019). Strikingly however, none displayed enrichment along the A/P axis (Fig. C.6.4E).

The posterior-enrichment of non-canonical Wnt ligands was especially interesting, and for multiple reasons. First, Wnt signaling is necessary to maintain stemness in human oviduct organoid cultures (Kessler *et al.* 2015), and secretory cells are thought to be the progenitors of MCCs in the mouse oviduct. Thus, it may be that posteriorly enriched Wnt signaling maintains the progenitor capabilities of posteriorly enriched secretory cells (Ghosh *et al.* 2017). Second, Wnt5a/7a/11 are known to orient PCP signaling, providing directional cues for the apical surface of cells and driving coordinated ciliary beating in multiciliated tissues (Gao *et al.* 2011; Ossipova *et al.* 2015; Butler and Wallingford 2017; Koyama *et al.* 2019). While the oviduct is planar polarized – i.e. Vangl2 localizes anteriorly in oviduct epithelial cells – it is unknown where the directional cue originates (Shi *et al.* 2016). Our data raise the possibility that posteriorly enriched Wnt5a/7a/11 orients Vangl2 localization to the anterior side of the oviduct epithelium, thereby regulating ciliary flow towards the posterior oviduct. Our experiments also indicate that PCP signaling components are evenly expressed across the A/P axis of the oviduct, which is consistent with the entire length of an epithelium being planar polarized. While the impact of planar polarization on MCCs is fairly well studied, the corresponding impact on secretory cells remains unknown.

C.5. CONCLUSION

In summary, our transcriptome profiling data fills a major gap in mouse oviduct investigations. While confirming reports made using other methods, our detailed anteroposterior axis and estrous stage analyses reveal novel gene expression patterns as well as providing a foundation for further studies. Hypotheses generated from these data can inform additional explorations using orthogonal methods such as single-cell sequencing and proteomics. Such studies have the potential to identify specific cell types associated with expression trends observed here, as well as quantifying actual cellular protein abundances which may differ from RNA expression.

C.6. FIGURES

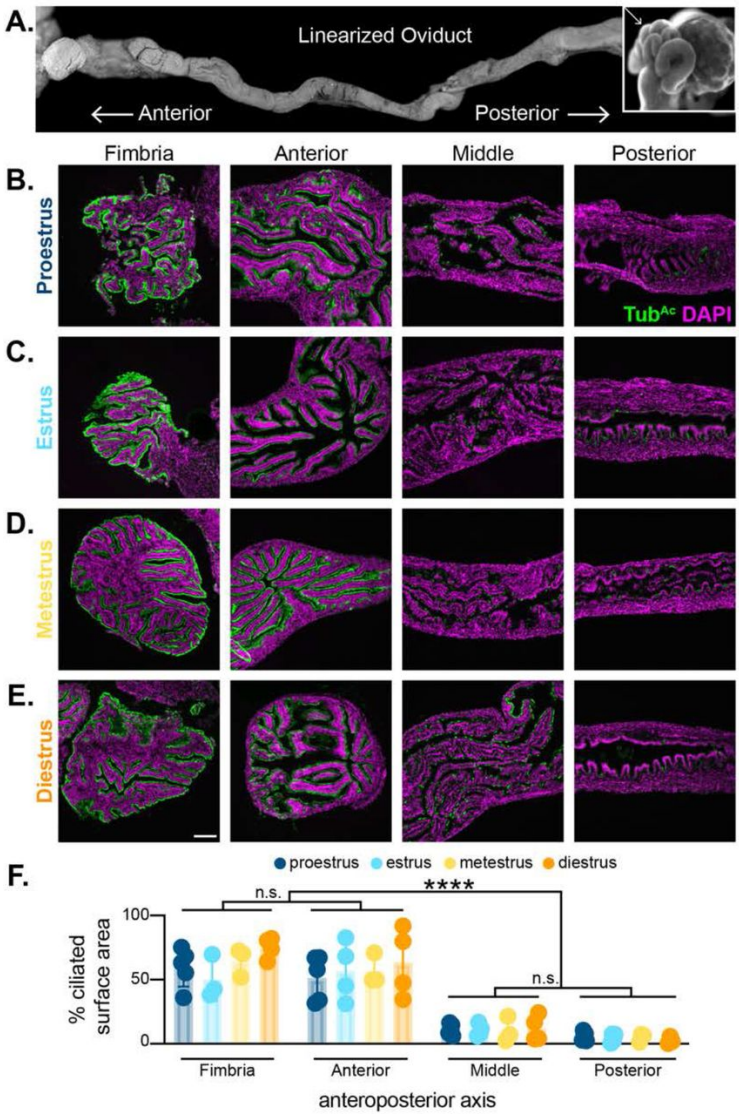


Figure C.6.1. The mouse oviduct displays anteroposterior patterning that does not change across the estrous cycle.

A) A linearized mouse oviduct, where the anterior is close to the ovary and the posterior is close to the uterus. The insert shows a non-linearized supercoiled mouse oviduct

(insert, arrow). Oviducts were collected at B) proestrus, C) estrus, D) metestrus, and E) diestrus, linearized, and imaged for nuclei (DAPI, magenta) and cilia (Tub^{Ac}, green) along the A/P axis (fimbria, anterior, middle, and posterior). Scale bar = 100μm. F) Quantitation of the percent ciliated surface area along the oviduct lumen in each of the four estrous stages (in temporal order: Proestrus, Estrus, Metestrus, Diestrus). Some error bars are too small to be seen. * = $p < 0.05$.

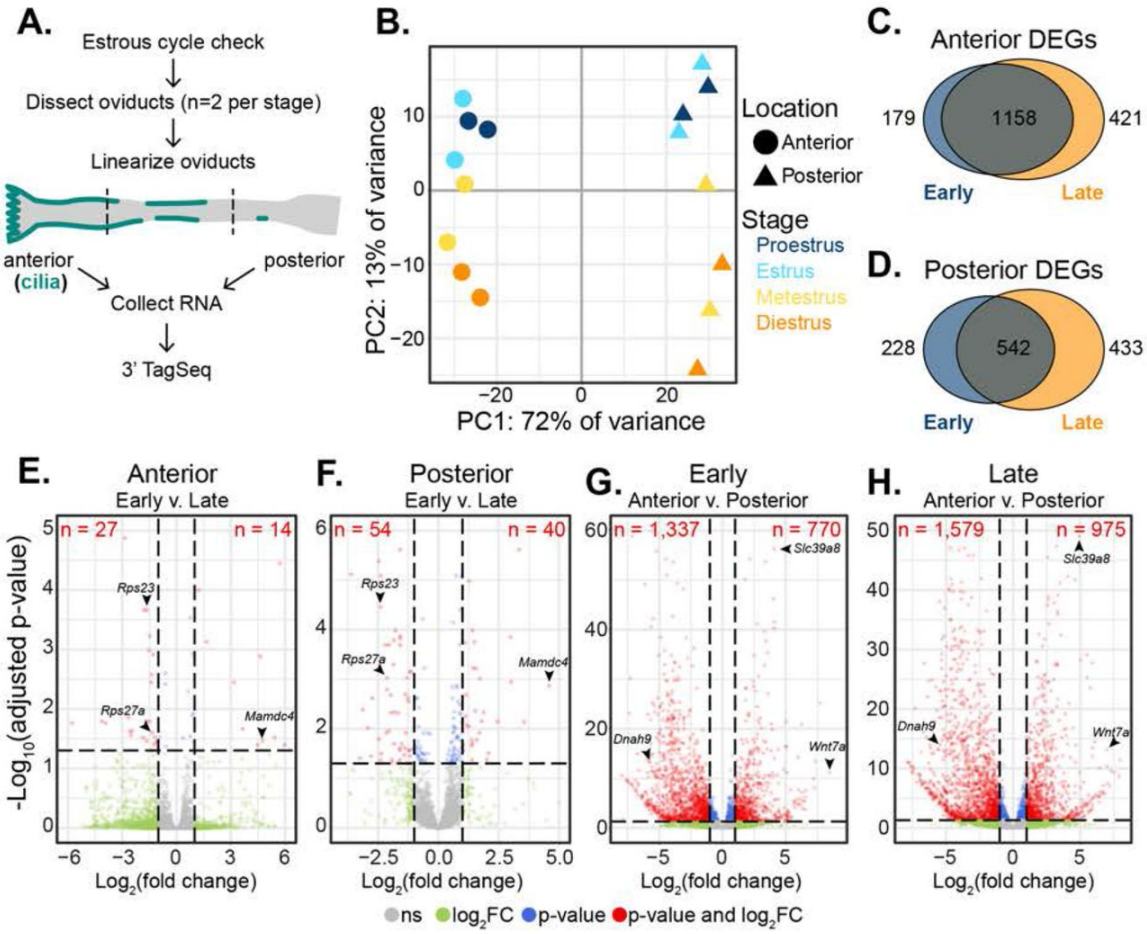


Figure C.6.2. RNAseq of the oviduct shows major differences along the anteroposterior axis and minor change across the estrous cycle.

A) Oviducts were dissected and linearized from each stage of the estrous cycle (n=2 each stage). RNA from the anterior and posterior thirds of each oviduct was collected and submitted for 3' TagSeq. B) Principal Component Analysis (PCA) of the sixteen 3'TagSeq datasets. A/P location (displayed as symbols) accounts for most of the variance while estrous cycle stage (displayed as colors) accounts for a lesser part with good separation between early (Proestrus, Estrus) and late (Metestrus, Diestrus) phases. C) Overlap between the anterior DEGs at early and late estrous cycle phases. D) Overlap

between the posterior DEGs at early and late phases of the estrous cycle. Volcano plots of oviduct E) anterior and F) posterior compare early and late DEGs. Volcano plots of estrous cycle G) early and H) late phases cycle show anterior and posterior DEGs. Red points are genes significant at both adjusted P-value (<= 0.05) and effect size (log2 fold change <= -1 or >= 1).

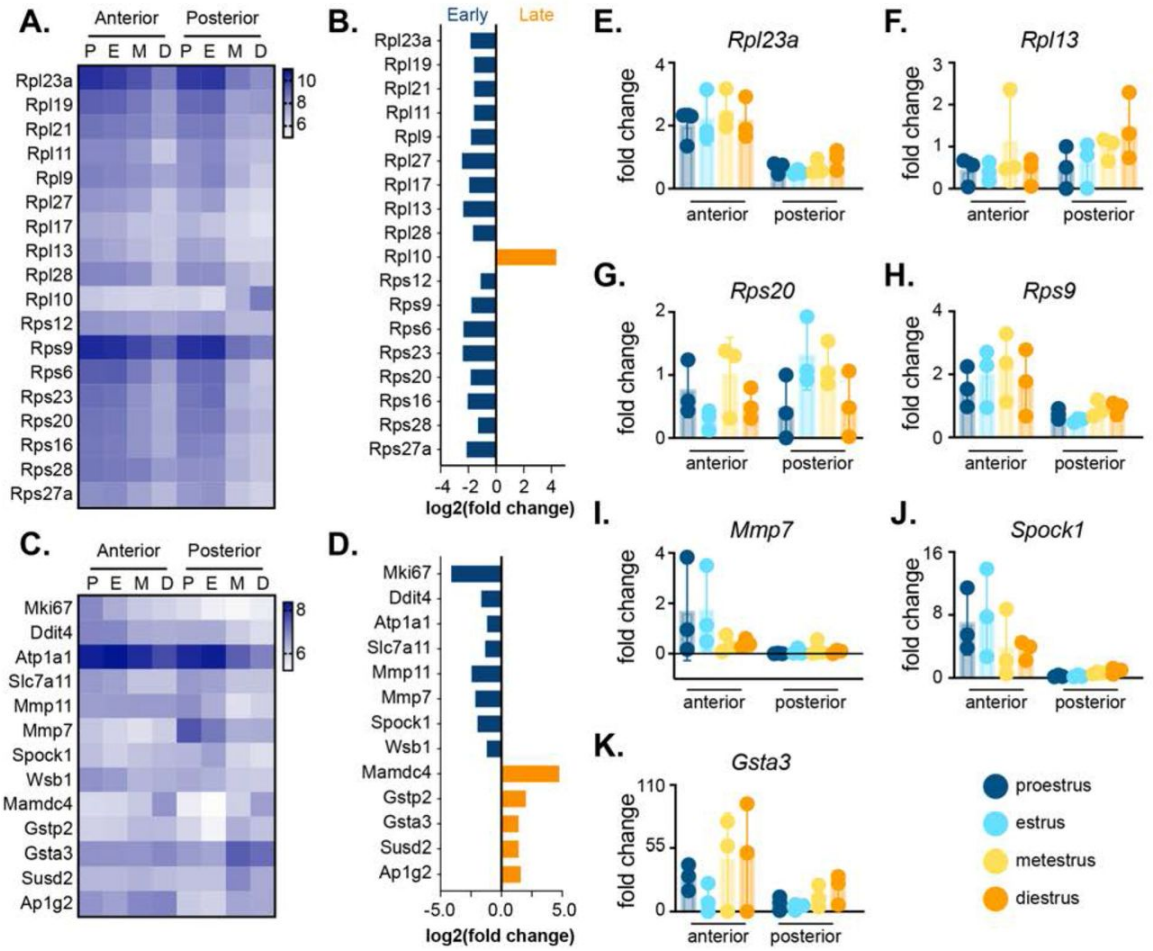


Figure C.6.3. The mouse oviduct transcriptome is remarkably stable across the estrous cycle.

A) Heatmap depicting expression of ribosomal genes across the estrous cycle where rows indicate genes and columns indicate the estrous cycle stage (P- proestrus, E- estrus, M- metestrus, D- diestrus). B) The effect size (absolute value of the log2 fold change) of the ribosomal genes was generally modest (under 2). C) Heatmap of other DEGs. D) These genes also generally have modest but consistent effect sizes. E-K) qPCR histograms of

select genes from B and D along the anteroposterior axis at each estrous cycle stage. All heatmap scales use DESeq2 variance stabilized counts.

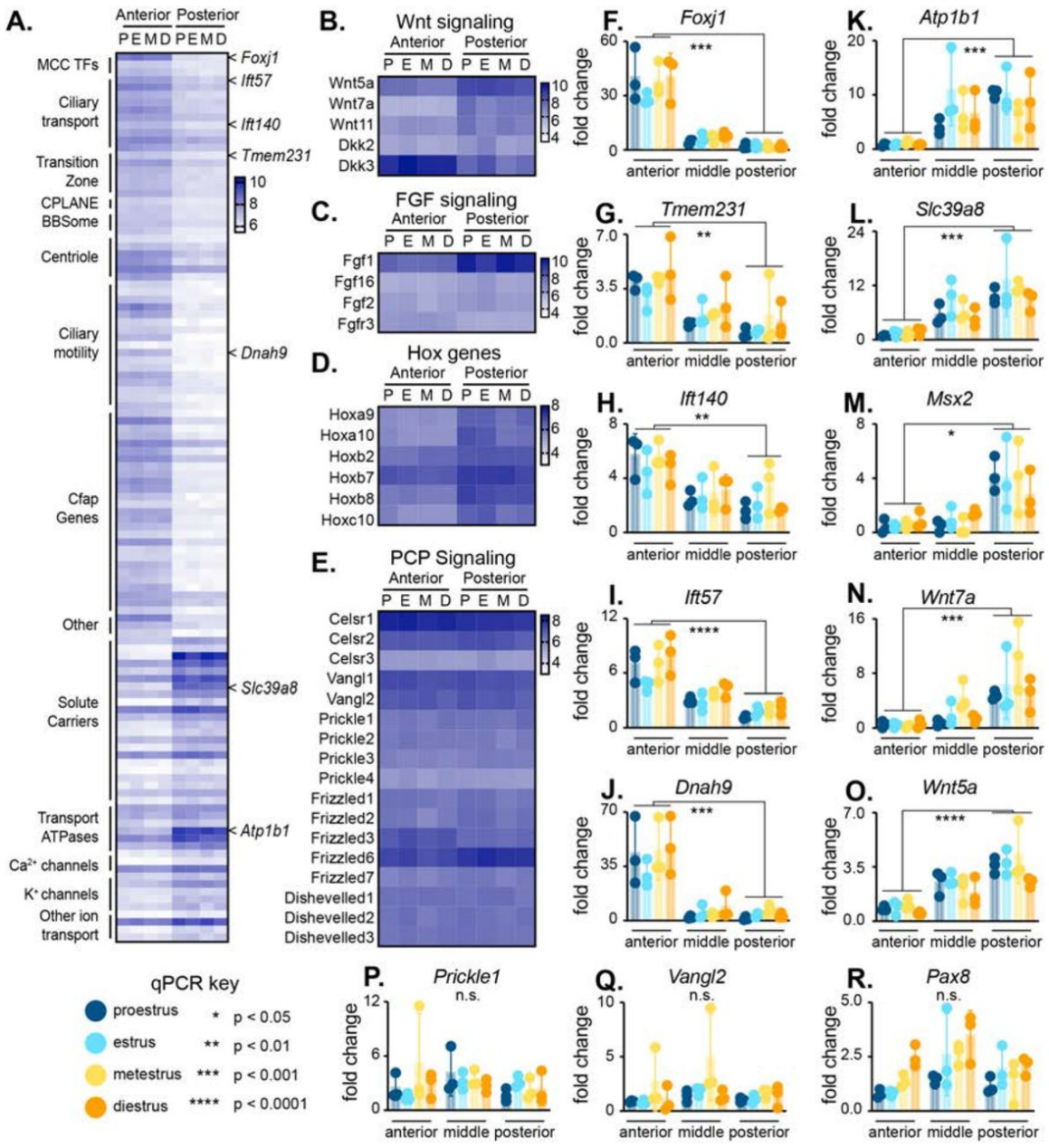


Figure C.6.4. The mouse oviduct displays robust transcriptional patterning along the anteroposterior axis.

A) Heatmap showing expression dynamics of DEGs across the anterior and posterior, including cilia-related genes and complexes, solute carriers, transport ATPases, and

calcium and potassium channels. Additional heatmaps of genes in developmental signaling pathways, including B) Wnt signaling, C) FGF signaling, D) Hox genes, and E) PCP signaling. F-R) qPCR histograms of select genes from A-E along the anteroposterior axis at each estrous cycle stage: anterior-enriched genes *Foxj1*, *Tmem231*, *Ift140*, *Ift57*, *Dnah9;* posterior-enriched genes *Atp1b1*, *Slc29a8*, *Msx2*, *Wnt7a*, *Wnt5a*; and other qPCR assayed genes *Prickle1*, *Vangl2*, *Pax8*.

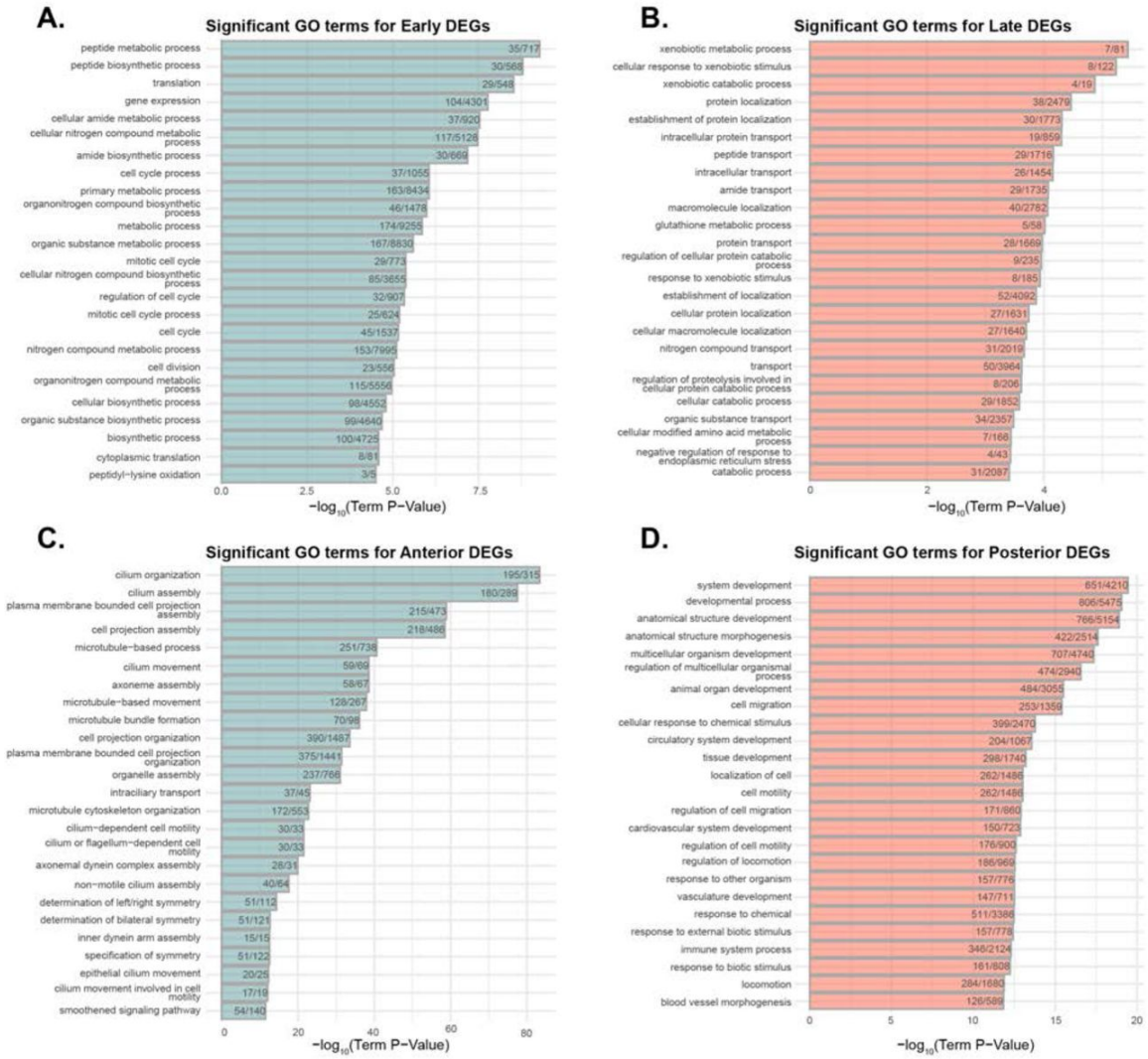


Figure C.6.S1. Significant GO terms associated with oviduct transcriptome analysis.

Histograms depicting significant GO terms for A) early, B) late, C) anterior, and D) posterior DEGs.

References

Abruzzi K. C., A. Smith, W. Chen, and F. Solomon, 2002 Protection from Free β-Tubulin by the β-Tubulin Binding Protein Rbl2p. Mol Cell Biol 22: 138–147. https://doi.org/10.1128/MCB.22.1.138-147.2002

Agapite J., L.-P. Albou, S. Aleksander, J. Argasinska, V. Arnaboldi, *et al.*, 2020 Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res 48: D650–D658. https://doi.org/10.1093/nar/gkz813

Agapite J., L.-P. Albou, S. Aleksander, J. Argasinska, V. Arnaboldi, *et al.*, Alliance of Genome Resources Portal: unified model organism research platform. Nucleic Acids Res. https://doi.org/10.1093/nar/gkz813

Agduhr E., 1927 Studies on the Structure and Development of the Bursa Ovarica and the Tuba Uterina in the Mouse. Acta Zoologica 8: 1–133. https://doi.org/10.1111/j.1463-6395.1927.tb00649.x

Agmon N., J. Temple, Z. Tang, T. Schraink, M. Baron, *et al.*, 2020 Phylogenetic debugging of a complete human biosynthetic pathway transplanted into yeast. Nucleic Acids Research 48: 486–499. https://doi.org/10.1093/nar/gkz1098

Aguayo-Ortiz R., O. Méndez-Lucio, J. L. Medina-Franco, R. Castillo, L. Yépez-Mulia, *et al.*, 2013 Towards the identification of the binding site of benzimidazoles to β-tubulin of Trichinella spiralis: Insights from computational and experimental data. Journal of Molecular Graphics and Modelling 41: 12–19. https://doi.org/10.1016/j.jmgm.2013.01.007

Aguayo-Ortiz R., O. Méndez-Lucio, R. Castillo, L. Yépez-Mulia, J. L. Medina-Franco, *et al.*, Molecular basis for benzimidazole resistance from a novel β-tubulin binding site model | Elsevier Enhanced Reader. Journal of Molecular Graphics and Modelling. https://doi.org/10.1016/j.jmgm.2013.07.008

Ahn E., P. Kumar, D. Mukha, A. Tzur, and T. Shlomi, 2017 Temporal fluxomics reveals oscillations in TCA cycle flux throughout the mammalian cell cycle. Molecular Systems Biology 13: 953. https://doi.org/10.15252/msb.20177763

Ajayi A. F., and R. E. Akhigbe, 2020 Staging of the estrous cycle and induction of estrus in experimental rodents: an update. Fertil Res Pract 6: 5. https://doi.org/10.1186/s40738-020-00074-3

Akhmetov A., J. Laurent, J. Gollihar, E. Gardner, R. Garge, *et al.*, 2018a Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. BIO-PROTOCOL 8. https://doi.org/10.21769/BioProtoc.2765

Akhmetov A., J. M. Laurent, J. Gollihar, E. C. Gardner, R. K. Garge, *et al.*, 2018b Single-step Precision Genome Editing in Yeast Using CRISPR-Cas9. Bio Protoc 8. https://doi.org/10.21769/BioProtoc.2765

Alberti S., A. D. Gitler, and S. Lindquist, 2007 A suite of Gateway® cloning vectors for high-throughput genetic analysis in Saccharomyces cerevisiae. Yeast 24: 913–919. https://doi.org/10.1002/yea.1502

Alexa A., and J. Rahnenfuhrer, 2020 *topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.*

Al-Feel W., J. C. DeMar, and S. J. Wakil, 2003 A Saccharomyces cerevisiae mutant strain defective in acetyl-CoA carboxylase arrests at the G2/M phase of the cell cycle. Proc Natl Acad Sci U S A 100: 3095–3100. https://doi.org/10.1073/pnas.0538069100

Ali Q., I. Rashid, M. Z. Shabbir, A.-U. Rahman, K. Shahzad, *et al.*, 2018 Emergence and the spread of the F200Y benzimidazole resistance mutation in Haemonchus contortus and Haemonchus placei from buffalo and cattle. bioRxiv 425660. https://doi.org/10.1101/425660

Amberg D. C., D. J. Burke, and J. N. Strathern, 2006 Yeast protein extracts. CSH Protoc 2006. https://doi.org/10.1101/pdb.prot4152

Amorosi C. J., M. A. Chiasson, M. G. McDonald, L. H. Wong, K. A. Sitko, *et al.*, 2021 Massively parallel characterization of CYP2C9 variant enzyme activity and abundance. bioRxiv 2021.03.12.435209. https://doi.org/10.1101/2021.03.12.435209

Aramayo R., and M. Polymenis, 2017 Ribosome profiling the cell cycle: lessons and challenges. Curr Genet 63: 959–964. https://doi.org/10.1007/s00294-017-0698-3

Atilla-Gokcumen G. E., E. Muro, J. Relat-Goberna, S. Sasse, A. Bedigian, *et al.*, 2014 Dividing cells regulate their lipid composition and localization. Cell 156: 428–439. https://doi.org/10.1016/j.cell.2013.12.015

Backman T. W. H., Y. Cao, and T. Girke, 2011 ChemMine tools: an online service for analyzing and clustering small molecules. Nucleic Acids Res 39: W486–W491. https://doi.org/10.1093/nar/gkr320

Balakrishnan R., J. Park, K. Karra, B. C. Hitz, G. Binkley, *et al.*, 2012 YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) 2012: bar062. https://doi.org/10.1093/database/bar062

Baltrušis P., P. Halvarsson, and J. Höglund, 2018 Exploring benzimidazole resistance in Haemonchus contortus by next generation sequencing and droplet digital PCR. International Journal for Parasitology: Drugs and Drug Resistance 8: 411–419. https://doi.org/10.1016/j.ijpddr.2018.09.003

Bandyopadhyay S., R. Sharan, and T. Ideker, 2006 Systematic identification of functional orthologs based on protein network comparison. Genome Res 16: 428–435. https://doi.org/10.1101/gr.4526006

Banfalvi G., 2008 Cell cycle synchronization of animal cells and nuclei by centrifugal elutriation. Nature Protocols 3: 663–673. https://doi.org/10.1038/nprot.2008.34

Banno S., F. Fukumori, A. Ichiishi, K. Okada, H. Uekusa, *et al.*, 2008 Genotyping of Benzimidazole-Resistant and Dicarboximide-Resistant Mutations in Botrytis cinerea Using Real-Time Polymerase Chain Reaction Assays. Phytopathology 98: 397–404. https://doi.org/10.1094/PHYTO-98-4-0397

Barbier P., A. Dorléans, F. Devred, L. Sanz, D. Allegro, *et al.*, 2010 Stathmin and Interfacial Microtubule Inhibitors Recognize a Naturally Curved Conformation of Tubulin Dimers. J. Biol. Chem. 285: 31672–31681. https://doi.org/10.1074/jbc.M110.141929

Bardwell L., 2005 A walk-through of the yeast mating pheromone response pathway. Peptides 26: 339–350.

Barger S. R., M. L. James, C. D. Pellenz, M. Krendel, and V. Sirotkin, 2019 Human myosin 1e tail but not motor domain replaces fission yeast Myo1 domains to support myosin-I function during endocytosis. Exp. Cell Res. 384: 111625. https://doi.org/10.1016/j.yexcr.2019.111625

Barton B., G. Herrera, P. Anamthathmakula, J. Rock, A. Willie, *et al.*, 2020 Roles of steroid hormones in oviductal function. Reproduction 159: R125–R137.

Bauersachs S., S. Rehfeld, S. E. Ulbrich, S. Mallok, K. Prelle, *et al.*, 2004 Monitoring gene expression changes in bovine oviduct epithelial cells during the oestrous cycle. J Mol Endocrinol 32: 449–466. https://doi.org/10.1677/jme.0.0320449

Becher I., A. Andrés-Pons, N. Romanov, F. Stein, M. Schramm, *et al.*, 2018 Pervasive Protein Thermal Stability Variation during the Cell Cycle. Cell 173: 1495-1507.e18. https://doi.org/10.1016/j.cell.2018.03.053

Beilharz T. H., P. F. Harrison, D. M. Miles, M. M. See, U. M. M. Le, *et al.*, 2017 Coordination of Cell Cycle Progression and Mitotic Spindle Assembly Involves Histone H3 Lysine 4 Methylation by Set1/COMPASS. Genetics 205: 185–199. https://doi.org/10.1534/genetics.116.194852

Binder J. X., S. Pletscher-Frankild, K. Tsafou, C. Stolte, S. I. O'Donoghue, *et al.*, 2014 COMPARTMENTS: unification and visualization of protein subcellular localization evidence. Database (Oxford) 2014: bau012. https://doi.org/10.1093/database/bau012

Blank H. M., R. Perez, C. He, N. Maitra, R. Metz, *et al.*, 2017 Translational control of lipogenic enzymes in the cell cycle of synchronous, growing yeast cells. EMBO J 36: 487–502. https://doi.org/10.15252/embj.201695050

Bode C. J., M. L. Gupta, K. A. Suprenant, and R. H. Himes, 2003 The two α-tubulin isotypes in budding yeast have opposing effects on microtubule dynamics in vitro. EMBO reports 4: 94–99. https://doi.org/10.1038/sj.embor.embor716

Boeke J. D., J. Trueheart, G. Natsoulis, and G. R. Fink, 1987 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol 154: 164–175. https://doi.org/10.1016/0076-6879(87)54076-9

Bray N. L., H. Pimentel, P. Melsted, and L. Pachter, 2016 Near-optimal probabilistic RNA-seq quantification. Nature Biotechnology 34: 525–527. https://doi.org/10.1038/nbt.3519

Brenner R. M., 1969 Renewal of oviduct cilia during the menstrual cycle of the rhesus monkey. Fertil Steril 20: 599–611. https://doi.org/10.1016/s0015-0282(16)37086-8

Brenner R. M., and N. B. West, 1975 Hormonal regulation of the reproductive tract in female mammals. Annu Rev Physiol 37: 273–302. https://doi.org/10.1146/annurev.ph.37.030175.001421

Bronson F. H., and T. H. Hamilton, 1970 *Concentrations of Nucleic Acids and Protein in the Mouse Oviduct During the Oestrous Cycle and After Steroid Replacement*.

Brown M. C., G. S. Taylor, and H. a. S. Epton, 1984 Carbendazim resistance in the eyespot pathogen Pseudocercosporella herpotrichoides. Plant Pathology 33: 101–111. https://doi.org/10.1111/j.1365-3059.1984.tb00593.x

Burga A., M. O. Casanueva, and B. Lehner, 2011 Predicting mutation outcome from early stochastic variation in genetic interaction partners. Nature 480: 250–253. https://doi.org/10.1038/nature10665

Burke D., P. Gasdaska, and L. Hartwell, 1989 Dominant effects of tubulin overexpression in Saccharomyces cerevisiae. Mol Cell Biol 9: 1049–1059.

Butler M. T., and J. B. Wallingford, 2017 Planar cell polarity in development and disease. Nature Reviews Molecular Cell Biology 18: 375–388. https://doi.org/10.1038/nrm.2017.11

Bylander A., M. Nutu, R. Wellander, M. Goksör, H. Billig, *et al.*, 2010 Rapid effects of progesterone on ciliary beat frequency in the mouse fallopian tube. Reprod Biol Endocrinol 8: 48. https://doi.org/10.1186/1477-7827-8-48

Cabañas R., G. Castellá, M. L. Abarca, M. R. Bragulat, and F. J. Cabañes, 2009 Thiabendazole resistance and mutations in the β-tubulin gene of Penicillium expansum strains isolated from apples and pears with blue mold decay. FEMS Microbiol Lett 297: 189–195. https://doi.org/10.1111/j.1574-6968.2009.01670.x

Cai S. X., 2007 Small molecule vascular disrupting agents: potential new drugs for cancer treatment. Recent Pat Anticancer Drug Discov 2: 79–101. https://doi.org/10.2174/157489207779561462

Cao Y., A. Charisi, L.-C. Cheng, T. Jiang, and T. Girke, 2008 ChemmineR: a compound mining framework for R. Bioinformatics 24: 1733–1734. https://doi.org/10.1093/bioinformatics/btn307

Carmeliet P., 2005 Angiogenesis in life, disease and medicine. Nature 438: 932–936. https://doi.org/10.1038/nature04478

Carpy A., K. Krug, S. Graf, A. Koch, S. Popic, *et al.*, 2014 Absolute proteome and phosphoproteome dynamics during the cell cycle of Schizosaccharomyces pombe (Fission Yeast). Mol Cell Proteomics 13: 1925–1936. https://doi.org/10.1074/mcp.M113.035824

Carter H. E., H. J. Cools, J. S. West, M. W. Shaw, and B. A. Fraaije, 2013 Detection and molecular characterisation of Pyrenopeziza brassicae isolates resistant to methyl benzimidazole carbamates. Pest Management Science 69: 1040–1048. https://doi.org/10.1002/ps.3585

Caudron F., E. Denarier, J.-C. Thibout-Quintana, J. Brocard, A. Andrieux, *et al.*, 2010 Mutation of Ser172 in yeast β tubulin induces defects in microtubule dynamics and cell division. PLoS ONE 5: e13553. https://doi.org/10.1371/journal.pone.0013553

Cerny K. L., E. Garrett, A. J. Walton, L. H. Anderson, and P. J. Bridges, 2015 A transcriptomal analysis of bovine oviductal epithelial cells collected during the follicular phase versus the luteal phase of the estrous cycle. Reprod Biol Endocrinol 13: 84. https://doi.org/10.1186/s12958-015-0077-1

Cha H. J., M. Byrom, P. E. Mead, A. D. Ellington, J. B. Wallingford, *et al.*, 2012 Evolutionarily Repurposed Networks Reveal the Well-Known Antifungal Drug Thiabendazole to Be a Novel Vascular Disrupting Agent. PLOS Biology 10: e1001379. https://doi.org/10.1371/journal.pbio.1001379

Chagas A. M., F. D. Sampaio Junior, A. Pacheco, A. B. da Cunha, J. dos S. Cruz, *et al.*, 2016 F200Y polymorphism of the β-tubulin isotype 1 gene in Haemonchus contortus and sheep flock management practices related to anthelmintic resistance in eastern Amazon. Veterinary Parasitology 226: 104–108. https://doi.org/10.1016/j.vetpar.2016.06.038

Chen X., H. Niu, W.-H. Chung, Z. Zhu, A. Papusha, *et al.*, 2011 Cell cycle regulation of DNA double-strand break end resection by Cdk1-dependent Dna2 phosphorylation. Nat Struct Mol Biol 18: 1015–1019. https://doi.org/10.1038/nsmb.2105

Chen B., W. Wang, X. Peng, H. Jiang, S. Zhang, *et al.*, 2019 The comprehensive mutational and phenotypic spectrum of TUBB8 in female infertility. European Journal of Human Genetics 27: 300. https://doi.org/10.1038/s41431-018-0283-3

Cherry J. M., C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, *et al.*, 1998 SGD: Saccharomyces Genome Database. Nucleic Acids Res 26: 73–79. https://doi.org/10.1093/nar/26.1.73

Cherry J. M., E. L. Hong, C. Amundsen, R. Balakrishnan, G. Binkley, *et al.*, 2012 Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res 40: D700–D705. https://doi.org/10.1093/nar/gkr1029

Cho R. J., M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, *et al.*, 1998 A genome-wide transcriptional analysis of the mitotic cell cycle. Mol Cell 2: 65–73. https://doi.org/10.1016/s1097-2765(00)80114-8

Christensen D. G., J. G. Meyer, J. T. Baumgartner, A. K. D'Souza, W. C. Nelson, *et al.*, 2018 Identification of Novel Protein Lysine Acetyltransferases in Escherichia coli. mBio 9. https://doi.org/10.1128/mBio.01905-18

Clark A. M., and P. Labute, 2007 2D Depiction of Protein−Ligand Complexes. J. Chem. Inf. Model. 47: 1933–1944. https://doi.org/10.1021/ci7001473

Collins B. C., C. L. Hunter, Y. Liu, B. Schilling, G. Rosenberger, *et al.*, 2017 Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nature Communications 8: 291. https://doi.org/10.1038/s41467-017-00249-5

Conant G. C., and K. H. Wolfe, 2008 Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet 9: 938–950. https://doi.org/10.1038/nrg2482

Coy P., F. García-Vázquez, P. Visconti, and M. Avilés, 2012 Roles of the oviduct in mammalian fertilization. Reproduction 144: 649–660. https://doi.org/10.1530/REP-12-0279

Coy P., and R. Yanagimachi, 2015 The Common and Species-Specific Roles of Oviductal Proteins in Mammalian Fertilization and Embryo Development. BioScience 65: 973–984. https://doi.org/10.1093/biosci/biv119

Creanor J., and J. M. Mitchison, 1979 Reduction of Perturbations in Leucine Incorporation in Synchronous Cultures of Schizosaccharomyces pombe Made by Elutriation. Microbiology, 112: 385–388. https://doi.org/10.1099/00221287-112-2-385

Csárdi G., A. Franks, D. S. Choi, E. M. Airoldi, and D. A. Drummond, 2015 Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS Genet 11: e1005206. https://doi.org/10.1371/journal.pgen.1005206

Curie S., 2019 Humanizing yeast sterol biosynthesis to understand evolution and disease

Dahl C., H. P. Biemann, and J. Dahl, 1987 A protein kinase antigenically related to pp60v-src possibly involved in yeast cell cycle control: positive in vivo regulation by sterol. Proc Natl Acad Sci U S A 84: 4012–4016. https://doi.org/10.1073/pnas.84.12.4012

Dai L., T. Zhao, X. Bisteau, W. Sun, N. Prabhu, *et al.*, 2018 Modulation of Protein-Interaction States through the Cell Cycle. Cell 173: 1481-1494.e13. https://doi.org/10.1016/j.cell.2018.03.065

Davidse L. C., and W. Flach, 1978 Interaction of thiabendazole with fungal tubulin. Biochimica et Biophysica Acta (BBA) - General Subjects 543: 82–90. https://doi.org/10.1016/0304-4165(78)90456-7

Davidse L. C., 1986 Benzimidazole Fungicides: Mechanism of Action and Biological Impact. Annual Review of Phytopathology 24: 43–65. https://doi.org/10.1146/annurev.py.24.090186.000355

Dawson P. J., W. E. Gutteridge, and K. Gull, 1984 A comparison of the interaction of anthelmintic benzimidazoles with tubulin isolated from mammalian tissue and the parasitic nematode Ascaridia galli. Biochemical Pharmacology 33: 1069–1074. https://doi.org/10.1016/0006-2952(84)90515-X

Dephoure N., C. Zhou, J. Villén, S. A. Beausoleil, C. E. Bakalarski, *et al.*, 2008 A quantitative atlas of mitotic phosphorylation. Proc Natl Acad Sci U S A 105: 10762–10767. https://doi.org/10.1073/pnas.0805139105

DiCarlo J. E., J. E. Norville, P. Mali, X. Rios, J. Aach, *et al.*, 2013 Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems. Nucl. Acids Res. 41: 4336–4343. https://doi.org/10.1093/nar/gkt135

Douglas L. M., F. J. Alvarez, C. McCreary, and J. B. Konopka, 2005 Septin Function in Yeast Model Systems and Pathogenic Fungi. Eukaryotic Cell 4: 1503–1512. https://doi.org/10.1128/EC.4.9.1503-1512.2005

Douzery E. J. P., E. A. Snell, E. Bapteste, F. Delsuc, and H. Philippe, 2004 The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? Proc Natl Acad Sci U S A 101: 15386–15391. https://doi.org/10.1073/pnas.0403984101

Driscoll M., E. Dean, E. Reilly, E. Bergholz, and M. Chalfie, 1989 Genetic and molecular analysis of a Caenorhabditis elegans beta-tubulin that conveys benzimidazole sensitivity. J. Cell Biol. 109: 2993–3003.

Du H., and H. S. Taylor, 2016 The Role of Hox Genes in Female Reproductive Tract Development, Adult Function, and Fertility. Cold Spring Harb Perspect Med 6. https://doi.org/10.1101/cshperspect.a023002

Duda D. M., R. C. A. M. van Waardenburg, L. A. Borg, S. McGarity, A. Nourse, *et al.*, 2007 Structure of a SUMO-binding-motif mimic bound to Smt3p-Ubc9p: conservation of a non-covalent ubiquitin-like protein-E2 complex as a platform for selective interactions within a SUMO pathway. J Mol Biol 369: 619–630. https://doi.org/10.1016/j.jmb.2007.04.007

Dunham M. J., and D. M. Fowler, 2013 Contemporary, yeast-based approaches to understanding human genetic variation. Current Opinion in Genetics & Development 23: 658–664. https://doi.org/10.1016/j.gde.2013.10.001

Dunham M. J., M. R. Gartenberg, and G. W. Brown (Eds.), 2015 *Methods in Yeast Genetics and Genomics: A Cold Spring Harbor Laboratory Course Manual, 2015 Edition*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Dunn C. W., F. Zapata, C. Munro, S. Siebert, and A. Hejnol, 2018 Pairwise comparisons across species are problematic when analyzing functional genomic data. Proc Natl Acad Sci U S A 115: E409–E417. https://doi.org/10.1073/pnas.1707515115

El-Kenawi A. E., and A. B. El-Remessy, 2013 Angiogenesis inhibitors in cancer therapy: mechanistic perspective on classification and treatment rationales. Br J Pharmacol 170: 712–729. https://doi.org/10.1111/bph.12344

Elledge S. J., and M. R. Spottswood, 1991 A new human p34 protein kinase, CDK2, identified by complementation of a cdc28 mutation in Saccharomyces cerevisiae, is a homolog of Xenopus Eg1. EMBO J 10: 2653–2659.

Elliott S. G., J. R. Warner, and C. S. McLaughlin, 1979 Synthesis of ribosomal proteins during the cell cycle of the yeast Saccharomyces cerevisiae. J Bacteriol 137: 1048–1050.

EPA, 2002 Thiabendazole and salts R.E.D Factsheet. United States Environmental Protection Agency.

Evans T., E. T. Rosenthal, J. Youngblom, D. Distel, and T. Hunt, 1983 Cyclin: a protein specified by maternal mRNA in sea urchin eggs that is destroyed at each cleavage division. Cell 33: 389–396. https://doi.org/10.1016/0092-8674(83)90420-8

Ewald J. C., A. Kuehne, N. Zamboni, and J. M. Skotheim, 2016 The Yeast Cyclin-Dependent Kinase Routes Carbon Fluxes to Fuel Cell Cycle Progression. Mol Cell 62: 532–545. https://doi.org/10.1016/j.molcel.2016.02.017

Ewels P., M. Magnusson, S. Lundin, and M. Käller, 2016 MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Falk J., L. Boubakar, and V. Castellani, 2019 Septin functions during neuro-development, a yeast perspective. Current Opinion in Neurobiology 57: 102–109. https://doi.org/10.1016/j.conb.2019.01.012

Feng R., Q. Sang, Y. Kuang, X. Sun, Z. Yan, *et al.*, 2016 Mutations in TUBB8 and Human Oocyte Meiotic Arrest. New England Journal of Medicine 374: 223–232. https://doi.org/10.1056/NEJMoa1510791

Ferenczy A., R. M. Richart, F. J. Agate, M. L. Purkerson, and E. W. Dempsey, 1972 Scanning electron microscopy of the human fallopian tube. Science 175: 783–784. https://doi.org/10.1126/science.175.4023.783

Flory M. R., H. Lee, R. Bonneau, P. Mallick, K. Serikawa, *et al.*, 2006 Quantitative proteomic analysis of the budding yeast cell cycle using acid-cleavable isotope-coded affinity tag reagents. Proteomics 6: 6146–6157. https://doi.org/10.1002/pmic.200600159

Folkman J., 2004 Endogenous angiogenesis inhibitors. APMIS 112: 496–507. https://doi.org/10.1111/j.1600-0463.2004.apm11207-0809.x

Folkman J., 2007 Opinion: Angiogenesis: an organizing principle for drug discovery? Nature Reviews Drug Discovery 6: 273–286. https://doi.org/10.1038/nrd2115

Fowler D. M., and S. Fields, 2014 Deep mutational scanning: a new style of protein science. Nat Meth 11: 801–807. https://doi.org/10.1038/nmeth.3027

Fowler D. M., J. J. Stephany, and S. Fields, 2014 Measuring the activity of protein variants on a large scale using deep mutational scanning. Nat. Protocols 9: 2267–2284. https://doi.org/10.1038/nprot.2014.153

Frankish A., M. Diekhans, A.-M. Ferreira, R. Johnson, I. Jungreis, *et al.*, 2019 GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47: D766–D773. https://doi.org/10.1093/nar/gky955

Frenkel-Morgenstern M., T. Danon, T. Christian, T. Igarashi, L. Cohen, *et al.*, 2012 Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol 8: 572. https://doi.org/10.1038/msb.2012.3

Fujishima K., H. Kiyonari, J. Kurisu, T. Hirano, and M. Kengaku, 2007 Targeted disruption of Sept3, a heteromeric assembly partner of Sept5 and Sept7 in axons, has no effect on developing CNS neurons. Journal of Neurochemistry 102: 77–92. https://doi.org/10.1111/j.1471-4159.2007.04478.x

Furse S., and G. C. Shearman, 2018 Do lipids shape the eukaryotic cell cycle? Biochim Biophys Acta Mol Cell Biol Lipids 1863: 9–19. https://doi.org/10.1016/j.bbalip.2017.09.010

Furtado L. F. V., A. C. P. de P. Bello, H. A. dos Santos, M. R. S. Carvalho, and É. M. L. Rabelo, 2014 First identification of the F200Y SNP in the β-tubulin gene linked to benzimidazole resistance in Ancylostoma caninum. Veterinary Parasitology 206: 313–316. https://doi.org/10.1016/j.vetpar.2014.10.021

Furtado L. F. V., and É. M. L. Rabelo, 2015 Molecular analysis of the F167Y SNP in the β-tubulin gene by screening genotypes of two Ancylostoma caninum populations. Veterinary Parasitology 210: 114–117. https://doi.org/10.1016/j.vetpar.2015.03.018

Gabaldón T., and E. V. Koonin, 2013 Functional and evolutionary implications of gene orthology. Nat Rev Genet 14: 360–366. https://doi.org/10.1038/nrg3456

Gaber R. F., D. M. Copple, B. K. Kennedy, M. Vidal, and M. Bard, 1989 The yeast gene ERG6 is required for normal membrane function but is not essential for biosynthesis of the cell-cycle-sparking sterol. Mol Cell Biol 9: 3447–3456. https://doi.org/10.1128/mcb.9.8.3447

Gao B., H. Song, K. Bishop, G. Elliot, L. Garrett, *et al.*, 2011 Wnt signaling gradients establish planar cell polarity by inducing Vangl2 phosphorylation through Ror2. Dev Cell 20: 163–176. https://doi.org/10.1016/j.devcel.2011.01.001

Garcia G., D. R. Raleigh, and J. F. Reiter, 2018 How the Ciliary Membrane Is Organized Inside-Out to Communicate Outside-In. Curr Biol 28: R421–R434. https://doi.org/10.1016/j.cub.2018.03.010

Garge R. K., J. M. Laurent, A. H. Kachroo, and E. M. Marcotte, 2020 Systematic Humanization of the Yeast Cytoskeleton Discerns Functionally Replaceable from Divergent Human Genes. Genetics 215: 1153–1169. https://doi.org/10.1534/genetics.120.303378

Gasch A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, *et al.*, 2000 Genomic expression programs in the response of yeast cells to environmental changes. Mol Biol Cell 11: 4241–4257. https://doi.org/10.1091/mbc.11.12.4241

Gentleman R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, *et al.*, 2004 Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5: R80. https://doi.org/10.1186/gb-2004-5-10-r80

Ghosh A., S. M. Syed, and P. S. Tanwar, 2017 In vivo genetic cell lineage tracing reveals that oviductal secretory cells self-renew and give rise to ciliated cells. Development 144: 3031–3041. https://doi.org/10.1242/dev.149989

Giaever G., A. M. Chu, L. Ni, C. Connelly, L. Riles, *et al.*, 2002 Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387–391. https://doi.org/10.1038/nature00935

Gillet L. C., P. Navarro, S. Tate, H. Röst, N. Selevsek, *et al.*, 2012 Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 11: O111.016717. https://doi.org/10.1074/mcp.O111.016717

Gonella-Diaza A. M., S. C. da Silva Andrade, M. Sponchiado, G. Pugliesi, F. S. Mesquita, *et al.*, 2017 Oviductal transcriptional profiling of a bovine fertility model by next-generation sequencing. Genomics Data 13: 27–29. https://doi.org/10.1016/j.gdata.2017.06.004

Gossen B. D., S. R. Rimmer, and J. D. Holley, 2001 First Report of Resistance to Benomyl Fungicide in Sclerotinia sclerotiorum. Plant Disease 85: 1206–1206. https://doi.org/10.1094/PDIS.2001.85.11.1206C

Granovskaia M. V., L. J. Jensen, M. E. Ritchie, J. Toedling, Y. Ning, *et al.*, 2010 High-resolution transcription atlas of the mitotic cell cycle in budding yeast. Genome Biology 11: R24. https://doi.org/10.1186/gb-2010-11-3-r24

Guo J., B. A. Bryan, and M. Polymenis, 2004 Nutrient-specific effects in the coordination of cell growth with cell division in continuous cultures of Saccharomyces cerevisiae. Arch Microbiol 182: 326–330. https://doi.org/10.1007/s00203-004-0704-2

Guo H., N. T. Ingolia, J. S. Weissman, and D. P. Bartel, 2010 Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466: 835–840. https://doi.org/10.1038/nature09267

Hahnel S. R., S. Zdraljevic, B. C. Rodriguez, Y. Zhao, P. T. McGrath, *et al.*, 2018 Extreme allelic heterogeneity at a Caenorhabditis elegans beta-tubulin locus explains natural resistance to benzimidazoles. PLOS Pathogens 14: e1007226. https://doi.org/10.1371/journal.ppat.1007226

Hall P. A., and S. H. Russell, The pathobiology of the septin gene family. The Journal of Pathology 204: 489–505. https://doi.org/10.1002/path.1654

Hamza A., E. Tammpere, M. Kofoed, C. Keong, J. Chiang, *et al.*, 2015 Complementation of Yeast Genes with Human Genes as an Experimental Platform for Functional Testing of Human Genetic Variants. Genetics 201: 1263–1274. https://doi.org/10.1534/genetics.115.181099

Hamza A., M. R. M. Driessen, E. Tammpere, N. J. O'Neil, and P. Hieter, 2020 Cross-Species Complementation of Nonessential Yeast Genes Establishes Platforms for Testing Inhibitors of Human Proteins. Genetics 214: 735–747. https://doi.org/10.1534/genetics.119.302971

Hardy C. D., and N. R. Cozzarelli, 2005 A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure. Mol Microbiol 57: 1636–1652. https://doi.org/10.1111/j.1365-2958.2005.04799.x

Hasslacher M., A. S. Ivessa, F. Paltauf, and S. D. Kohlwein, 1993 Acetyl-CoA carboxylase from yeast is an essential enzyme and is regulated by factors that control phospholipid metabolism. J Biol Chem 268: 10946–10952.

Hauser A. S., S. Chavali, I. Masuho, L. J. Jahn, K. A. Martemyanov, *et al.*, 2018 Pharmacogenomics of GPCR Drug Targets. Cell 172: 41-54.e19. https://doi.org/10.1016/j.cell.2017.11.033

Heath V. L., and R. Bicknell, 2009 Anticancer strategies involving the vasculature. Nature Reviews Clinical Oncology 6: 395–404. https://doi.org/10.1038/nrclinonc.2009.52

Heinicke S., M. S. Livstone, C. Lu, R. Oughtred, F. Kang, *et al.*, 2007 The Princeton Protein Orthology Database (P-POD): A Comparative Genomics Analysis Tool for Biologists. PLOS ONE 2: e766. https://doi.org/10.1371/journal.pone.0000766

Herbert S. P., and D. Y. R. Stainier, 2011 Molecular control of endothelial cell behaviour during blood vessel morphogenesis. Nature Reviews Molecular Cell Biology 12: 551–564. https://doi.org/10.1038/nrm3176

Herrera G. G. B., S. L. Lierz, E. A. Harris, L. J. Donoghue, S. C. Hewitt, *et al.*, 2020 Oviductal Retention of Embryos in Female Mice Lacking Estrogen Receptor α in the Isthmus and the Uterus. Endocrinology 161. https://doi.org/10.1210/endocr/bqz033

Hess A. P., S. Talbi, A. E. Hamilton, D. M. Baston-Buest, M. Nyegaard, *et al.*, 2013 The human oviduct transcriptome reveals an anti-inflammatory, anti-angiogenic, secretory and matrix-stable environment during embryo transit. Reprod Biomed Online 27: 423–435. https://doi.org/10.1016/j.rbmo.2013.06.013

Hinnen P., and F. a. L. M. Eskens, 2007 Vascular disrupting agents in clinical development. British Journal of Cancer 96: 1159–1165. https://doi.org/10.1038/sj.bjc.6603694

Ho B., A. Baryshnikova, and G. W. Brown, 2018 Unification of Protein Abundance Datasets Yields a Quantitative Saccharomyces cerevisiae Proteome. Cell Syst 6: 192-205.e3. https://doi.org/10.1016/j.cels.2017.12.004

Hoose S. A., J. A. Rawlings, M. M. Kelly, M. C. Leitch, Q. O. Ababneh, *et al.*, 2012 A systematic analysis of cell cycle regulators in yeast reveals that most factors act independently of cell size to control initiation of division. PLoS Genet 8: e1002590. https://doi.org/10.1371/journal.pgen.1002590

Hopper A. K., 2013 Transfer RNA post-transcriptional processing, turnover, and subcellular dynamics in the yeast Saccharomyces cerevisiae. Genetics 194: 43–67. https://doi.org/10.1534/genetics.112.147470

Huang J., C. J. Mousley, L. Dacquay, N. Maitra, G. Drin, *et al.*, 2018a A Lipid Transfer Protein Signaling Axis Exerts Dual Control of Cell-Cycle and Membrane Trafficking Systems. Dev Cell 44: 378-391.e5. https://doi.org/10.1016/j.devcel.2017.12.026

Huang H., T. Yang, Q. Shao, T. Majumder, K. Mell, *et al.*, 2018b Human TUBB3 Mutations Disrupt Netrin Attractive Signaling. Neuroscience 374: 155–171. https://doi.org/10.1016/j.neuroscience.2018.01.046

Huerta-Cepas J., D. Szklarczyk, K. Forslund, H. Cook, D. Heller, *et al.*, 2016 eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 44: D286–D293. https://doi.org/10.1093/nar/gkv1248

Huh W.-K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, *et al.*, 2003 Global analysis of protein localization in budding yeast. Nature 425: 686–691. https://doi.org/10.1038/nature02026

Ibrahim M. A., D. V. Do, Y. J. Sepah, S. M. Shah, E. Van Anden, *et al.*, 2013 Vascular disrupting agent for neovascular age related macular degeneration: a pilot study of the safety and efficacy of intravenous combretastatin a-4 phosphate. BMC Pharmacology and Toxicology 14: 7. https://doi.org/10.1186/2050-6511-14-7

Ingolia N. T., S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman, 2009 Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science 324: 218–223. https://doi.org/10.1126/science.1168978

Innan H., and F. Kondrashov, 2010 The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet 11: 97–108. https://doi.org/10.1038/nrg2689

Jambre L. F. L., P. J. Martin, and R. F. Webb, 1979 Thiabendazole Resistance in Field Populations of Haemonchus Contortus. Australian Veterinary Journal 55: 163–166. https://doi.org/10.1111/j.1751-0813.1979.tb15263.x

Janke C., 2014 The tubulin code: Molecular components, readout mechanisms, and functions. J Cell Biol 206: 461–472. https://doi.org/10.1083/jcb.201406055

Jékely G., 2014 Origin and Evolution of the Self-Organizing Cytoskeleton in the Network of Eukaryotic Organelles. Cold Spring Harb Perspect Biol 6: a016030. https://doi.org/10.1101/cshperspect.a016030

Jüppner J., U. Mubeen, A. Leisse, C. Caldana, H. Brust, *et al.*, 2017 Dynamics of lipids and metabolites during the cell cycle of Chlamydomonas reinhardtii. Plant J 92: 331–343. https://doi.org/10.1111/tpj.13642

Kachroo A. H., J. M. Laurent, C. M. Yellman, A. G. Meyer, C. O. Wilke, *et al.*, 2015 Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science 348: 921–925. https://doi.org/10.1126/science.aaa0769

Kachroo A. H., J. M. Laurent, A. Akhmetov, M. Szilagyi-Jones, C. D. McWhite, *et al.*, 2017 Systematic bacterialization of yeast genes identifies a near-universally swappable pathway. eLife 6: e25093. https://doi.org/10.7554/eLife.25093

Kaplan C., M. Steinmann, N. A. Zapiorkowska, and H. Ewers, 2017 Functional Redundancy of Septin Homologs in Dendritic Branching. Front. Cell Dev. Biol. 5. https://doi.org/10.3389/fcell.2017.00011

Kataoka T., S. Powers, S. Cameron, O. Fasano, M. Goldfarb, *et al.*, 1985 Functional homology of mammalian and yeast RAS genes. Cell 40: 19–26. https://doi.org/10.1016/0092-8674(85)90304-6

Katoh K., and D. M. Standley, 2013 MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30: 772–780. https://doi.org/10.1093/molbev/mst010

Kavallaris M., 2010 Microtubules and resistance to tubulin-binding agents. Nat Rev Cancer 10: 194–204. https://doi.org/10.1038/nrc2803

Kearse M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, *et al.*, 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keaton M. A., and D. J. Lew, 2006 Eavesdropping on the cytoskeleton: progress and controversy in the yeast morphogenesis checkpoint. Current Opinion in Microbiology 9: 540–546. https://doi.org/10.1016/j.mib.2006.10.004

Keegan J. D., B. Good, T. de Waal, J. Fanning, and O. M. Keane, 2017 Genetic basis of benzimidazole resistance in Teladorsagia circumcincta in Ireland. Irish Veterinary Journal 70: 8. https://doi.org/10.1186/s13620-017-0087-8

Kent W. J., 2002 BLAT--the BLAST-like alignment tool. Genome Res 12: 656–664. https://doi.org/10.1101/gr.229202

Kerbel R. S., 2008 Tumor Angiogenesis. New England Journal of Medicine 358: 2039–2049. https://doi.org/10.1056/NEJMra0706596

Kessler M., K. Hoffmann, V. Brinkmann, O. Thieck, S. Jackisch, *et al.*, 2015 The Notch and Wnt pathways regulate stemness and differentiation in human fallopian tube organoids. Nat Commun 6: 8989. https://doi.org/10.1038/ncomms9989

Kettle E., S. L. Page, G. P. Morgan, C. S. Malladi, C. L. Wong, *et al.*, 2015 A Cholesterol-Dependent Endocytic Mechanism Generates Midbody Tubules During Cytokinesis. Traffic 16: 1174–1192. https://doi.org/10.1111/tra.12328

Kim Y., E.-Y. Kim, Y.-M. Seo, T. K. Yoon, W.-S. Lee, *et al.*, 2012 Function of the pentose phosphate pathway and its key enzyme, transketolase, in the regulation of the meiotic cell cycle in oocytes. Clin Exp Reprod Med 39: 58–67. https://doi.org/10.5653/cerm.2012.39.2.58

Kim J., and M. D. Rose, 2015 Stable Pseudohyphal Growth in Budding Yeast Induced by Synergism between Septin Defects and Altered MAP-kinase Signaling. PLOS Genetics 11: e1005684. https://doi.org/10.1371/journal.pgen.1005684

Kim S., P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, *et al.*, 2016 PubChem Substance and Compound databases. Nucleic Acids Res 44: D1202–D1213. https://doi.org/10.1093/nar/gkv951

Kim J.-M., J.-E. Park, I. Yoo, J. Han, N. Kim, *et al.*, 2018 Integrated transcriptomes throughout swine oestrous cycle reveal dynamic changes in reproductive tissues interacting networks. Scientific Reports 8: 5436. https://doi.org/10.1038/s41598-018-23655-1

Kofoed M., K. L. Milbury, J. H. Chiang, S. Sinha, S. Ben-Aroya, *et al.*, 2015 An Updated Collection of Sequence Barcoded Temperature-Sensitive Alleles of Yeast Essential Genes. G3 5: 1879–1887. https://doi.org/10.1534/g3.115.019174

Koonin E. V., 2005 Orthologs, Paralogs, and Evolutionary Genomics. Annu. Rev. Genet. 39: 309–338. https://doi.org/10.1146/annurev.genet.39.073003.114725

Koyama H., D. Shi, and T. Fujimori, 2019 Biophysics in oviduct: Planar cell polarity, cilia, epithelial fold and tube morphogenesis, egg dynamics. Biophys Physicobiol 16: 89–107. https://doi.org/10.2142/biophysico.16.0_89

Kulak N. A., G. Pichler, I. Paron, N. Nagaraj, and M. Mann, 2014 Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. Nat Methods 11: 319–324. https://doi.org/10.1038/nmeth.2834

Kumar S., R. Garg, S. Kumar, P. S. Banerjee, H. Ram, *et al.*, 2016 Benzimidazole resistance in equine cyathostomins in India. Veterinary Parasitology 218: 93–97. https://doi.org/10.1016/j.vetpar.2016.01.016

Kumar S., G. Stecher, M. Suleski, and S. B. Hedges, 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. Mol Biol Evol 34: 1812–1819. https://doi.org/10.1093/molbev/msx116

Kurat C. F., H. Wolinski, J. Petschnigg, S. Kaluarachchi, B. Andrews, *et al.*, 2009 Cdk1/Cdc28-dependent activation of the major triacylglycerol lipase Tgl4 in yeast links lipolysis to cell-cycle progression. Mol Cell 33: 53–63. https://doi.org/10.1016/j.molcel.2008.12.019

Kuscu C., S. Arslan, R. Singh, J. Thorpe, and M. Adli, 2014 Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nat Biotechnol 32: 677–683. https://doi.org/10.1038/nbt.2916

Labute P., 2009 Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. Proteins 75: 187–205. https://doi.org/10.1002/prot.22234

Labute P., 2010 LowModeMD—Implicit Low-Mode Velocity Filtering Applied to Conformational Search of Macrocycles and Protein Loops. J. Chem. Inf. Model. 50: 792–800. https://doi.org/10.1021/ci900508k

Lacey E., and J. H. Gill, 1994 Biochemistry of benzimidazole resistance. Acta Tropica 56: 245–262. https://doi.org/10.1016/0001-706X(94)90066-3

Lahtvee P.-J., B. J. Sánchez, A. Smialowska, S. Kasvandik, I. E. Elsemman, *et al.*, 2017 Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. Cell Syst 4: 495-504.e5. https://doi.org/10.1016/j.cels.2017.03.003

Lamesch P., N. Li, S. Milstein, C. Fan, T. Hao, *et al.*, 2007 hORFeome v3.1: A resource of human open reading frames representing over 10,000 human genes. Genomics 89: 307–315. https://doi.org/10.1016/j.ygeno.2006.11.012

Lane K. R., Y. Yu, P. E. Lackey, X. Chen, W. F. Marzluff, *et al.*, 2013 Cell cycle-regulated protein abundance changes in synchronously proliferating HeLa cells include regulation of pre-mRNA splicing proteins. PLoS One 8: e58456. https://doi.org/10.1371/journal.pone.0058456

Larsson O., N. Sonenberg, and R. Nadon, 2011 anota: Analysis of differential translation in genome-wide studies. Bioinformatics 27: 1440–1441. https://doi.org/10.1093/bioinformatics/btr146

Laurent J. M., J. H. Young, A. H. Kachroo, and E. M. Marcotte, 2016 Efforts to make and apply humanized yeast. Brief Funct Genomics 15: 155–163. https://doi.org/10.1093/bfgp/elv041

Laurent J. M., 2016 Evolutionary conservation of protein abundance and function

Laurent J. M., R. K. Garge, A. I. Teufel, C. O. Wilke, A. H. Kachroo, *et al.*, 2019 Humanization of yeast genes with multiple human orthologs reveals principles of functional divergence between paralogs. bioRxiv 668335. https://doi.org/10.1101/668335

Laurent J. M., R. K. Garge, A. I. Teufel, C. O. Wilke, A. H. Kachroo, *et al.*, 2020 Humanization of yeast genes with multiple human orthologs reveals functional divergence between paralogs. PLOS Biology 18: e3000627. https://doi.org/10.1371/journal.pbio.3000627

Leandro-García L. J., S. Leskelä, I. Landa, C. Montero-Conde, E. López-Jiménez, *et al.*, 2010 Tumoral and tissue-specific expression of the major human beta-tubulin isotypes. Cytoskeleton (Hoboken) 67: 214–223. https://doi.org/10.1002/cm.20436

Leaver-Fay A., M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, *et al.*, 2011 ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol 487: 545–574. https://doi.org/10.1016/B978-0-12-381270-4.00019-6

Lee M. G., and P. Nurse, 1987 Complementation used to clone a human homologue of the fission yeast cell cycle control gene cdc2. Nature 327: 31–35. https://doi.org/10.1038/327031a0

Lee I., U. M. Blom, P. I. Wang, J. E. Shim, and E. M. Marcotte, 2011 Prioritizing candidate disease genes by network-based boosting of genome-wide association data. Genome Res 21: 1109–1121. https://doi.org/10.1101/gr.118992.110

Lee M. E., W. C. DeLoache, B. Cervantes, and J. E. Dueber, 2015 A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth. Biol. 4: 975–986. https://doi.org/10.1021/sb500366v

Legendre M., L.-E. Zaragosi, and H. M. Mitchison, 2020 Motile cilia and airway disease. Seminars in Cell & Developmental Biology. https://doi.org/10.1016/j.semcdb.2020.11.007

Lewis M., and T. H. Stracker, 2020 Transcriptional regulation of multiciliated cell differentiation. Semin Cell Dev Biol. https://doi.org/10.1016/j.semcdb.2020.04.007

Li Z., F. J. Vizeacoumar, S. Bahr, J. Li, J. Warringer, *et al.*, 2011 Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat Biotech 29: 361–367. https://doi.org/10.1038/nbt.1832

Liang X., L. Peng, C.-H. Baek, and F. Katzen, 2013 Single step BP/LR combined Gateway reactions. Biotechniques 55: 265–268. https://doi.org/10.2144/000114101

Liao Y., G. K. Smyth, and W. Shi, 2019 The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res 47: e47. https://doi.org/10.1093/nar/gkz114

Lichtenberg U. de, R. Wernersson, T. S. Jensen, H. B. Nielsen, A. Fausbøll, *et al.*, 2005 New weakly expressed cell cycle-regulated genes in yeast. Yeast 22: 1191–1201. https://doi.org/10.1002/yea.1302

Lin Y., H. Liu, Z. Liu, Y. Liu, Q. He, *et al.*, 2013 Development and evaluation of an entirely solution-based combinative sample preparation method for membrane proteomics. Anal Biochem 432: 41–48. https://doi.org/10.1016/j.ab.2012.09.023

Lindahl P. E., 1948 Principle of a Counter-streaming Centrifuge for the Separation of Particles of Different Sizes. Nature 161: 648–649. https://doi.org/10.1038/161648a0

Lindsey R., Y. Ha, and M. Momany, 2010 A Septin from the Filamentous Fungus A. nidulans Induces Atypical Pseudohyphae in the Budding Yeast S. cerevisiae. PLOS ONE 5: e9858. https://doi.org/10.1371/journal.pone.0009858

Lippert J. W., 2007 Vascular disrupting agents. Bioorganic & Medicinal Chemistry 15: 605–615. https://doi.org/10.1016/j.bmc.2006.10.020

Liu S., Z. Che, and G. Chen, 2016 Multiple-fungicide resistance to carbendazim, diethofencarb, procymidone, and pyrimethanil in field isolates of Botrytis cinerea from tomato in Henan Province, China. Crop Protection 84: 56–61. https://doi.org/10.1016/j.cropro.2016.02.012

Liu S., Y. Zhang, J. Jiang, Z. Che, Y. Tian, *et al.*, 2018 Carbendazim resistance and dimethachlone sensitivity of field isolates of Sclerotinia sclerotiorum from oilseed rape in Henan Province, China. Journal of Phytopathology 166: 701–708. https://doi.org/10.1111/jph.12751

Locey K. J., and J. T. Lennon, 2016 Scaling laws predict global microbial diversity. Proc Natl Acad Sci U S A 113: 5970–5975. https://doi.org/10.1073/pnas.1521291113

Lodish H. F., 1974 Model for the regulation of mRNA translation applied to haemoglobin synthesis. Nature 251: 385–388. https://doi.org/10.1038/251385a0

Lohman B. K., J. N. Weber, and D. I. Bolnick, 2016 Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. Mol Ecol Resour 16: 1315–1321. https://doi.org/10.1111/1755-0998.12529

Love M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550. https://doi.org/10.1186/s13059-014-0550-8

Lu P., C. Vogel, R. Wang, X. Yao, and E. M. Marcotte, 2007 Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. Nat Biotechnol 25: 117–124. https://doi.org/10.1038/nbt1270

Lubega G. W., and R. K. Prichard, 1990 Specific interaction of benzimidazole anthelmintics with tubulin: high-affinity binding and benzimidazole resistance in Haemonchus contortus. Molecular and Biochemical Parasitology 38: 221–232. https://doi.org/10.1016/0166-6851(90)90025-H

Luongo T. S., J. M. Eller, M.-J. Lu, M. Niere, F. Raith, *et al.*, 2020 SLC25A51 is a mammalian mitochondrial NAD + transporter. Nature 588: 174–179. https://doi.org/10.1038/s41586-020-2741-7

Ly T., A. Endo, and A. I. Lamond, 2015 Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells, (J. Pines, Ed.). eLife 4: e04534. https://doi.org/10.7554/eLife.04534

Lynch M., and J. S. Conery, 2000 The evolutionary fate and consequences of duplicate genes. Science 290: 1151–1155. https://doi.org/10.1126/science.290.5494.1151

Maillo V., C. de Frutos, P. O'Gaora, N. Forde, G. W. Burns, *et al.*, 2016 Spatial differences in gene expression in the bovine oviduct. Reproduction 152: 37–46. https://doi.org/10.1530/REP-16-0074

Maitra N., J. Anandhakumar, H. M. Blank, C. D. Kaplan, and M. Polymenis, 2019 Perturbations of Transcription and Gene Expression-Associated Processes Alter Distribution of Cell Size Values in Saccharomyces cerevisiae. G3 (Bethesda) 9: 239–250. https://doi.org/10.1534/g3.118.200854

Marzo M. G., J. M. Griswold, K. M. Ruff, R. E. Buchmeier, C. P. Fees, *et al.*, 2019 Molecular basis for dyneinopathies reveals insight into dynein regulation and dysfunction, (A. P. Carter, A. Akhmanova, R. J. McKenney, and G. Schiavo, Eds.). eLife 8: e47246. https://doi.org/10.7554/eLife.47246

Mason R. P., D. Zhao, L. Liu, M. L. Trawick, and K. G. Pinney, 2011 A Perspective on Vascular Disrupting Agents that Interact with Tubulin: Preclinical Tumor Imaging and Biological Assessment. Integr Biol (Camb) 3: 375–387. https://doi.org/10.1039/c0ib00135j

McCusker D., and D. R. Kellogg, 2012 Plasma membrane growth during the cell cycle: unsolved mysteries and recent progress. Curr Opin Cell Biol 24: 845–851. https://doi.org/10.1016/j.ceb.2012.10.008

McGary K. L., I. Lee, and E. M. Marcotte, 2007 Broad network-based predictability of Saccharomyces cerevisiae gene loss-of-function phenotypes. Genome Biol 8: R258. https://doi.org/10.1186/gb-2007-8-12-r258

McGary K. L., T. J. Park, J. O. Woods, H. J. Cha, J. B. Wallingford, *et al.*, 2010 Systematic discovery of nonobvious human disease models through orthologous phenotypes. PNAS 107: 6544–6549. https://doi.org/10.1073/pnas.0910200107

McKean P. G., S. Vaughan, and K. Gull, 2001 The extended tubulin superfamily. Journal of Cell Science 114: 2723–2733.

McMurray M. A., and J. Thorner, 2008 Septin Stability and Recycling during Dynamic Structural Transitions in Cell Division and Development. Current Biology 18: 1203–1208. https://doi.org/10.1016/j.cub.2008.07.020

McMurray M. A., A. Bertin, G. Garcia, L. Lam, E. Nogales, *et al.*, 2011 Septin Filament Formation Is Essential in Budding Yeast. Developmental Cell 20: 540–549. https://doi.org/10.1016/j.devcel.2011.02.004

Meyer E., G. V. Aglyamova, and M. V. Matz, 2011 Profiling gene expression responses of coral larvae (Acropora millepora) to elevated temperature and settlement inducers using a novel RNA-Seq procedure. Mol Ecol 20: 3599–3616. https://doi.org/10.1111/j.1365-294X.2011.05205.x

Michel A. H., R. Hatakeyama, P. Kimmig, M. Arter, M. Peter, *et al.*, 2017 Functional mapping of yeast genomes by saturated transposition, (R. J. Deshaies, Ed.). eLife 6: e23570. https://doi.org/10.7554/eLife.23570

Middelberg A., and P. B. McKenna, 1983 Oxfendazole resistance in Nematodirus spathiger. New Zealand Veterinary Journal 31: 65–66. https://doi.org/10.1080/00480169.1983.34971

Mimitou E. P., and L. S. Symington, 2009 DNA end resection: many nucleases make light work. DNA Repair (Amst) 8: 983–995. https://doi.org/10.1016/j.dnarep.2009.04.017

Molk J. N., and K. Bloom, 2006 Microtubule dynamics in the budding yeast mating pathway. J Cell Sci 119: 3485–3490. https://doi.org/10.1242/jcs.03193

Mushegian A. R., and E. V. Koonin, 1996 A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc Natl Acad Sci U S A 93: 10268–10273. https://doi.org/10.1073/pnas.93.19.10268

Nakaune R., and M. Nakano, 2007 Benomyl resistance of Colletotrichum acutatum is caused by enhanced expression of β-tubulin 1 gene regulated by putative leucine zipper protein CaBEN1. Fungal Genetics and Biology 44: 1324–1335. https://doi.org/10.1016/j.fgb.2007.03.007

Neff N. F., J. H. Thomas, P. Grisafi, and D. Botstein, 1983 Isolation of the β-tubulin gene from yeast and demonstration of its essential function in vivo. Cell 33: 211–219. https://doi.org/10.1016/0092-8674(83)90350-1

Nehrt N. L., W. T. Clark, P. Radivojac, and M. W. Hahn, 2011 Testing the ortholog conjecture with comparative functional genomic data from mammals. PLoS Comput Biol 7: e1002073. https://doi.org/10.1371/journal.pcbi.1002073

Neiman A. M., 2011 Sporulation in the Budding Yeast Saccharomyces cerevisiae. Genetics 189: 737–765. https://doi.org/10.1534/genetics.111.127126

Niciura S. C. M., C. J. Veríssimo, J. G. G. Gromboni, M. I. P. Rocha, S. S. de Mello, *et al.*, 2012 F200Y polymorphism in the β-tubulin gene in field isolates of Haemonchus contortus and risk factors of sheep flock management practices related to anthelmintic resistance. Veterinary Parasitology 190: 608–612. https://doi.org/10.1016/j.vetpar.2012.07.016

Nogales E., V. H. Ramey, and H.-W. Wang, 2010 Chapter 8 - Cryo-EM Studies of Microtubule Structural Intermediates and Kinetochore–Microtubule Interactions, pp. 128–156 in *Methods in Cell Biology*, Microtubules, in vitro. edited by Correia L. W. and J. J. Academic Press.

Novak E., and H. S. Everett, 1928 Cyclical and other variations in the tubal epithelium. American Journal of Obstetrics & Gynecology 16: 499–530. https://doi.org/10.1016/S0002-9378(28)90535-9

Nowak-Sliwinska P., H. van den Bergh, M. Sickenberg, and A. H. C. Koh, 2013 Photodynamic therapy for polypoidal choroidal vasculopathy. Progress in Retinal and Eye Research 37: 182–199. https://doi.org/10.1016/j.preteyeres.2013.09.003

Nyberg P., L. Xie, and R. Kalluri, 2005 Endogenous Inhibitors of Angiogenesis. Cancer Res 65: 3967–3979. https://doi.org/10.1158/0008-5472.CAN-04-2427

O'Boyle N. M., M. Banck, C. A. James, C. Morley, T. Vandermeersch, *et al.*, 2011 Open Babel: An open chemical toolbox. Journal of Cheminformatics 3: 33. https://doi.org/10.1186/1758-2946-3-33

Ohno S., 1970 *Evolution by gene duplication*. Springer.

Oliva A., A. Rosebrock, F. Ferrezuelo, S. Pyne, H. Chen, *et al.*, 2005 The Cell Cycle–Regulated Genes of Schizosaccharomyces pombe. PLOS Biology 3: e225. https://doi.org/10.1371/journal.pbio.0030225

Olsen J. V., M. Vermeulen, A. Santamaria, C. Kumar, M. L. Miller, *et al.*, 2010 Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3: ra3. https://doi.org/10.1126/scisignal.2000475

Olshen A. B., A. C. Hsieh, C. R. Stumpf, R. A. Olshen, D. Ruggero, *et al.*, 2013 Assessing gene-level translational control from ribosome profiling. Bioinformatics 29: 2995–3002. https://doi.org/10.1093/bioinformatics/btt533

O'Neil N. J., M. L. Bailey, and P. Hieter, 2017 Synthetic lethality and cancer. Nature Reviews Genetics 18: 613–623. https://doi.org/10.1038/nrg.2017.47

O'Reilly M. S., L. Holmgren, Y. Shing, C. Chen, R. A. Rosenthal, *et al.*, 1994 Angiostatin: a novel angiogenesis inhibitor that mediates the suppression of metastases by a Lewis lung carcinoma. Cell 79: 315–328. https://doi.org/10.1016/0092-8674(94)90200-3

O'Reilly M. S., T. Boehm, Y. Shing, N. Fukai, G. Vasios, *et al.*, 1997 Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88: 277–285. https://doi.org/10.1016/s0092-8674(00)81848-6

Ossipova O., K. Kim, and S. Y. Sokol, 2015 Planar polarization of Vangl2 in the vertebrate neural plate is controlled by Wnt and Myosin II signaling. Biol Open 4: 722–730. https://doi.org/10.1242/bio.201511676

Pan X., D. S. Yuan, D. Xiang, X. Wang, S. Sookhai-Mahadeo, *et al.*, 2004 A robust toolkit for functional profiling of the yeast genome. Mol Cell 16: 487–496. https://doi.org/10.1016/j.molcel.2004.09.035

Pan F., R. L. Malmberg, and M. Momany, 2007 Analysis of septins across kingdoms reveals orthology and new motifs. BMC Evol Biol 7: 103. https://doi.org/10.1186/1471-2148-7-103

Patir A., A. M. Fraser, M. W. Barnett, L. McTeir, J. Rainger, *et al.*, 2020 The transcriptional signature associated with human motile cilia. Sci Rep 10: 10814. https://doi.org/10.1038/s41598-020-66453-4

Peng J., D. Schwartz, J. E. Elias, C. C. Thoreen, D. Cheng, *et al.*, 2003 A proteomics approach to understanding protein ubiquitination. Nature Biotechnology 21: 921–926. https://doi.org/10.1038/nbt849

Pramila T., W. Wu, S. Miles, W. S. Noble, and L. L. Breeden, 2006 The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle. Genes Dev 20: 2266–2278. https://doi.org/10.1101/gad.1450606

Ramünke S., L. Melville, L. Rinaldi, H. Hertzberg, T. de Waal, *et al.*, 2016 Benzimidazole resistance survey for Haemonchus, Teladorsagia and Trichostrongylus in three European countries using pyrosequencing including the development of new assays for Trichostrongylus. International Journal for Parasitology: Drugs and Drug Resistance 6: 230–240. https://doi.org/10.1016/j.ijpddr.2016.10.002

Ranaivoson F. M., B. Gigant, S. Berritt, M. Joullié, and M. Knossow, 2012 Structural plasticity of tubulin assembly probed by vinca-domain ligands. Acta Crystallogr D Biol Crystallogr 68: 927–934. https://doi.org/10.1107/S0907444912017143

Remm M., C. E. Storm, and E. L. Sonnhammer, 2001 Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314: 1041–1052. https://doi.org/10.1006/jmbi.2000.5197

Roberson E. C., N. K. Tran, M. J. Konjikusic, R. D. Fitch, R. S. Gray, *et al.*, 2020 A comparative study of the turnover of multiciliated cells in the mouse trachea, oviduct, and brain. Dev Dyn 249: 898–905. https://doi.org/10.1002/dvdy.165

Rohl C. A., C. E. M. Strauss, D. Chivian, and D. Baker, 2004 Modeling structurally variable regions in homologous proteins with rosetta. Proteins 55: 656–677. https://doi.org/10.1002/prot.10629

Romero R. A., and T. B. Sutton, 1998 Characterization of Benomyl Resistance in Mycosphaerella fijiensis, Cause of Black Sigatoka of Banana, in Costa Rica. Plant Disease 82: 931–934. https://doi.org/10.1094/PDIS.1998.82.8.931

Rose I. M., M. Bidarimath, A. Webster, A. K. Godwin, A. Flesken-Nikitin, *et al.*, 2020 WNT and inflammatory signaling distinguish human Fallopian tube epithelial cell populations. Scientific Reports 10: 9837. https://doi.org/10.1038/s41598-020-66556-y

Rual J.-F., K. Venkatesan, T. Hao, T. Hirozane-Kishikawa, A. Dricot, *et al.*, 2005 Towards a proteome-scale map of the human protein–protein interaction network. Nature 437: 1173–1178. https://doi.org/10.1038/nature04209

Rupp S., C. Plesken, S. Rumsey, M. Dowling, G. Schnabel, *et al.*, 2017 Botrytis fragariae, a New Species Causing Gray Mold on Strawberries, Shows High Frequencies of Specific and Efflux-Based Fungicide Resistance. Appl. Environ. Microbiol. 83. https://doi.org/10.1128/AEM.00269-17

Rustici G., J. Mata, K. Kivinen, P. Lió, C. J. Penkett, *et al.*, 2004 Periodic gene expression program of the fission yeast cell cycle. Nat Genet 36: 809–817. https://doi.org/10.1038/ng1377

Saeed M., Z. Iqbal, and A. Jabbar, 2007 Oxfendazole Resistance in Gastrointestinal Nematodes of Beetal Goats at Livestock Farms of Punjab (Pakistan). Acta Vet. Brno 76: 79–85. https://doi.org/10.2754/avb200776010079

Saito T. L., M. Ohtani, H. Sawai, F. Sano, A. Saka, *et al.*, 2004 SCMD: Saccharomyces cerevisiae Morphological Database. Nucleic Acids Res. 32: D319-322. https://doi.org/10.1093/nar/gkh113

Sanchez-Alvarez M., Q. Zhang, F. Finger, M. J. O. Wakelam, and C. Bakal, 2015 Cell cycle progression is an essential regulatory component of phospholipid metabolism and membrane homeostasis. Open Biol 5: 150093. https://doi.org/10.1098/rsob.150093

Santos A., R. Wernersson, and L. J. Jensen, 2015 Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. Nucleic Acids Res 43: D1140-1144. https://doi.org/10.1093/nar/gku1092

Santos J. M. L. dos, J. F. Vasconcelos, G. A. Frota, W. L. C. Ribeiro, W. P. P. André, *et al.*, 2017 Haemonchus contortus β-tubulin isotype 1 gene F200Y and F167Y SNPs are both selected by ivermectin and oxfendazole treatments with differing impacts on anthelmintic resistance. Veterinary Parasitology 248: 90–95. https://doi.org/10.1016/j.vetpar.2017.11.003

Scaglia N., S. Tyekucheva, G. Zadra, C. Photopoulos, and M. Loda, 2014 De novo fatty acid synthesis at the mitotic exit is required to complete cellular division. Cell Cycle 13: 859–868. https://doi.org/10.4161/cc.27767

Schatz P. J., L. Pillus, P. Grisafi, F. Solomon, and D. Botstein, 1986a Two functional alpha-tubulin genes of the yeast Saccharomyces cerevisiae encode divergent proteins. Mol. Cell. Biol. 6: 3711–3721. https://doi.org/10.1128/MCB.6.11.3711

Schatz P. J., F. Solomon, and D. Botstein, 1986b Genetically essential and nonessential alpha-tubulin genes specify functionally interchangeable proteins. Mol. Cell. Biol. 6: 3722–3733. https://doi.org/10.1128/MCB.6.11.3722

Schilling B., B. W. Gibson, and C. L. Hunter, 2017 Generation of High-Quality SWATH® Acquisition Data for Label-free Quantitative Proteomics Studies Using TripleTOF® Mass Spectrometers. Methods Mol Biol 1550: 223–233. https://doi.org/10.1007/978-1-4939-6747-6_16

Schillinger J., K. Severin, F. Kaschani, M. Kaiser, and M. Ehrmann, 2018 HTRA1-Dependent Cell Cycle Proteomics. J Proteome Res 17: 2679–2694. https://doi.org/10.1021/acs.jproteome.8b00129

Schindelin J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, *et al.*, 2012 Fiji: an open-source platform for biological-image analysis. Nature Methods 9: 676–682. https://doi.org/10.1038/nmeth.2019

Schindelin J., C. T. Rueden, M. C. Hiner, and K. W. Eliceiri, 2015 The ImageJ ecosystem: An open platform for biomedical image analysis. Molecular Reproduction and Development 82: 518–529. https://doi.org/10.1002/mrd.22489

Schneiter R., M. Hitomi, A. S. Ivessa, E. V. Fasch, S. D. Kohlwein, *et al.*, 1996 A yeast acetyl coenzyme A carboxylase mutant links very-long-chain fatty acid synthesis to the structure and function of the nuclear membrane-pore complex. Mol Cell Biol 16: 7161–7172. https://doi.org/10.1128/mcb.16.12.7161

Sd A., N. Dl, T. V, M. M, M. T, *et al.*, 2012 A link between mitotic entry and membrane growth suggests a novel model for cell size control. J Cell Biol 197: 89–104. https://doi.org/10.1083/jcb.201108108

Shi D., K. Komatsu, T. Uemura, and T. Fujimori, 2011 Analysis of ciliary beat frequency and ovum transport ability in the mouse oviduct. Genes Cells 16: 282–290. https://doi.org/10.1111/j.1365-2443.2011.01484.x

Shi D., K. Komatsu, M. Hirao, Y. Toyooka, H. Koyama, *et al.*, 2014 Celsr1 is required for the generation of polarity at multiple levels of the mouse oviduct. Development 141: 4558–4568. https://doi.org/10.1242/dev.115659

Shi D., F. Usami, K. Komatsu, S. Oka, T. Abe, *et al.*, 2016 Dynamics of planar cell polarity protein Vangl2 in the mouse oviduct epithelium. Mech Dev 141: 78–89. https://doi.org/10.1016/j.mod.2016.05.002

Shirley B., and R. L. Reeder, 1996 Cyclic changes in the ampulla of the rat oviduct. J Exp Zool 276: 164–173. https://doi.org/10.1002/(SICI)1097-010X(19961001)276:2<164::AID-JEZ10>3.0.CO;2-K

Shulman R. W., L. H. Hartwell, and J. R. Warner, 1973 Synthesis of ribosomal proteins during the yeast cell cycle. J Mol Biol 73: 513–525. https://doi.org/10.1016/0022-2836(73)90097-1

Singh P., R. Saxena, G. Srinivas, G. Pande, and A. Chattopadhyay, 2013 Cholesterol biosynthesis and homeostasis in regulation of the cell cycle. PLoS One 8: e58833. https://doi.org/10.1371/journal.pone.0058833

Sirajuddin M., L. M. Rice, and R. D. Vale, 2014 Regulation of microtubule motors by tubulin isotypes and post-translational modifications. Nat. Cell Biol. 16: 335–344. https://doi.org/10.1038/ncb2920

Sisu C., P. Muir, A. Frankish, I. Fiddes, M. Diekhans, *et al.*, 2020 Transcriptional activity and strain-specific history of mouse pseudogenes. Nature Communications 11: 3695. https://doi.org/10.1038/s41467-020-17157-w

Skuce P., L. Stenhouse, F. Jackson, V. Hypša, and J. Gilleard, 2010 Benzimidazole resistance allele haplotype diversity in United Kingdom isolates of Teladorsagia circumcincta supports a hypothesis of multiple origins of resistance by recurrent mutation. International Journal for Parasitology 40: 1247–1255. https://doi.org/10.1016/j.ijpara.2010.03.016

Soifer I., and N. Barkai, 2014 Systematic identification of cell size regulators in budding yeast. Mol Syst Biol 10. https://doi.org/10.15252/msb.20145345

Soma S., K. Yang, M. I. Morales, and M. Polymenis, 2014 Multiple metabolic requirements for size homeostasis and initiation of division in Saccharomyces cerevisiae. Microb Cell 1: 256–266. https://doi.org/10.15698/mic2014.08.160

Soneson C., M. I. Love, and M. D. Robinson, 2015 Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 4: 1521. https://doi.org/10.12688/f1000research.7563.2

Sonnhammer E. L. L., and G. Östlund, 2015 InParanoid 8: orthology analysis between 273 proteomes, mostly eukaryotic. Nucleic Acids Res 43: D234–D239. https://doi.org/10.1093/nar/gku1203

Soukup S. W., Evolution by gene duplication. S. Ohno. Springer-Verlag, New York. 1970. 160 pp. Teratology 9: 250–251. https://doi.org/10.1002/tera.1420090224

Sowamber R., O. Nelson, L. Dodds, V. DeCastro, I. Paudel, *et al.*, 2020 Integrative Transcriptome Analyses of the Human Fallopian Tube: Fimbria and Ampulla-Site of Origin of Serous Carcinoma of the Ovary. Cancers (Basel) 12. https://doi.org/10.3390/cancers12051090

Spellman P. T., G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, *et al.*, 1998 Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol Biol Cell 9: 3273–3297. https://doi.org/10.1091/mbc.9.12.3273

Stamatakis A., 2014 RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stark C., B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz, *et al.*, 2006 BioGRID: a general repository for interaction datasets. Nucleic Acids Res 34: D535-539. https://doi.org/10.1093/nar/gkj109

Stepanova T., J. Slemmer, C. C. Hoogenraad, G. Lansbergen, B. Dortland, *et al.*, 2003 Visualization of Microtubule Growth in Cultured Neurons via the Use of EB3-GFP (End-Binding Protein 3-Green Fluorescent Protein). J. Neurosci. 23: 2655–2664. https://doi.org/10.1523/JNEUROSCI.23-07-02655.2003

Stewart C. A., and R. R. Behringer, 2012 Mouse oviduct development. Results Probl Cell Differ 55: 247–262. https://doi.org/10.1007/978-3-642-30406-4_14

Stottmann R. W., A. Driver, A. Gutierrez, M. R. Skelton, M. Muntifering, *et al.*, 2016 A Heterozygous Mutation in Tubulin, Beta 2b (tubb2b) Causes Cognitive Deficits and Hippocampal Disorganization. Genes, Brain and Behavior n/a-n/a. https://doi.org/10.1111/gbb.12327

Studer R. A., and M. Robinson-Rechavi, 2009 How confident can we be that orthologs are similar, but paralogs differ? Trends Genet 25: 210–216. https://doi.org/10.1016/j.tig.2009.03.004

Suárez Y., C. Fernández, B. Ledo, A. J. Ferruelo, M. Martín, *et al.*, 2002 Differential effects of ergosterol and cholesterol on Cdk1 activation and SRE-driven transcription. Eur J Biochem 269: 1761–1771. https://doi.org/10.1046/j.1432-1327.2002.02822.x

Suarez S. S., 2016 Mammalian sperm interactions with the female reproductive tract. Cell Tissue Res 363: 185–194. https://doi.org/10.1007/s00441-015-2244-2

Sun S., F. Yang, G. Tan, M. Costanzo, R. Oughtred, *et al.*, 2016 An extended set of yeast-based functional assays accurately identifies human disease mutations. Genome Res. 26: 670–680. https://doi.org/10.1101/gr.192526.115

Sun S., J. Weile, M. Verby, A. G. Cote, Y. Wu, *et al.*, 2018 A proactive genotype-to-patient-phenotype map for cystathionine beta-synthase. bioRxiv 473983. https://doi.org/10.1101/473983

Swaffer M. P., A. W. Jones, H. R. Flynn, A. P. Snijders, and P. Nurse, 2016 CDK Substrate Phosphorylation and Ordering the Cell Cycle. Cell 167: 1750-1761.e16. https://doi.org/10.1016/j.cell.2016.11.034

Swaney D. L., P. Beltrao, L. Starita, A. Guo, J. Rush, *et al.*, 2013 Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. Nat Methods 10: 676–682. https://doi.org/10.1038/nmeth.2519

Talbot P., C. Geiske, and M. Knoll, 1999 Oocyte pickup by the mammalian oviduct. Mol Biol Cell 10: 5–8. https://doi.org/10.1091/mbc.10.1.5

Taylor R. D., M. MacCoss, and A. D. G. Lawson, 2014 Rings in Drugs: Miniperspective. J. Med. Chem. 57: 5845–5859. https://doi.org/10.1021/jm4017625

Temple G., D. S. Gerhard, R. Rasooly, E. A. Feingold, P. J. Good, *et al.*, 2009 The completion of the Mammalian Gene Collection (MGC). Genome Res. 19: 2324–2333. https://doi.org/10.1101/gr.095976.109

Teufel A. I., and C. O. Wilke, 2017 Accelerated simulation of evolutionary trajectories in origin-fixation models. J R Soc Interface 14. https://doi.org/10.1098/rsif.2016.0906

Teufel A. I., M. M. Johnson, J. M. Laurent, A. H. Kachroo, E. M. Marcotte, *et al.*, 2019 The Many Nuanced Evolutionary Consequences of Duplicated Genes. Mol Biol Evol 36: 304–314. https://doi.org/10.1093/molbev/msy210

The UniProt Consortium, 2019 UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res 47: D506–D515. https://doi.org/10.1093/nar/gky1049

Tischfield M. A., H. N. Baris, C. Wu, G. Rudolph, L. Van Maldergem, *et al.*, 2010 Human TUBB3 Mutations Perturb Microtubule Dynamics, Kinesin Interactions, and Axon Guidance. Cell 140: 74–87. https://doi.org/10.1016/j.cell.2009.12.011

Tolliver S. C., E. T. Lyons, J. H. Drudge, S. Stamper, and D. E. Granstrom, 1993 Critical tests of thiabendazole, oxibendazole, and oxfendazole for drug resistance of population-B equine small strongyles (1989 and 1990). Am. J. Vet. Res. 54: 908–913.

Tong A. H., M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, *et al.*, 2001 Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368. https://doi.org/10.1126/science.1065810

Torrent M., G. Chalancon, N. S. de Groot, A. Wuster, and M. Madan Babu, 2018 Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci Signal 11. https://doi.org/10.1126/scisignal.aat6409

Tozer G. M., C. Kanthou, and B. C. Baguley, 2005 Disrupting tumour blood vessels. Nat Rev Cancer 5: 423–435. https://doi.org/10.1038/nrc1628

Truong S. K., R. F. McCormick, and M. Polymenis, 2013 Genetic Determinants of Cell Size at Birth and Their Impact on Cell Cycle Progression in Saccharomyces cerevisiae. G3: Genes, Genomes, Genetics 3: 1525–1530. https://doi.org/10.1534/g3.113.007062

Tsang C. W., M. Fedchyshyn, J. Harrison, H. Xie, J. Xue, *et al.*, 2008 Superfluous Role of Mammalian Septins 3 and 5 in Neuronal Development and Synaptic Transmission. Molecular and Cellular Biology 28: 7012–7029. https://doi.org/10.1128/MCB.00035-08

Tumu S., A. Patil, W. Towns, M. Dyavaiah, and T. J. Begley, 2012 The gene-specific codon counting database: a genome-based catalog of one-, two-, three-, four- and five-codon combinations present in Saccharomyces cerevisiae genes. Database (Oxford) 2012: bas002. https://doi.org/10.1093/database/bas002

Uhlén M., L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, *et al.*, 2015 Tissue-based map of the human proteome. Science 347: 1260419. https://doi.org/10.1126/science.1260419

Vela-Corcía D., D. Romero, A. de Vicente, and A. Pérez-García, 2018 Analysis of β-tubulin-carbendazim interaction reveals that binding site for MBC fungicides does not include residues involved in fungicide resistance. Scientific Reports 8: 7161. https://doi.org/10.1038/s41598-018-25336-5

Verhage H. G., J. H. Abel, W. J. Tietz, and M. D. Barrau, 1973 Development and maintenance of the oviductal epithelium during the estrous cycle in the bitch. Biol Reprod 9: 460–474. https://doi.org/10.1093/biolreprod/9.5.460

Vogel J., B. Drapkin, J. Oomen, D. Beach, K. Bloom, *et al.*, 2001 Phosphorylation of γ-Tubulin Regulates Microtubule Organization in Budding Yeast. Developmental Cell 1: 621–631. https://doi.org/10.1016/S1534-5807(01)00073-9

Vogel C., and E. M. Marcotte, 2012 Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet 13: 227–232. https://doi.org/10.1038/nrg3185

Wach A., A. Brachat, R. Pöhlmann, and P. Philippsen, 1994 New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. Yeast 10: 1793–1808.

Wan C., B. Borgeson, S. Phanse, F. Tu, K. Drew, *et al.*, 2015 Panorama of ancient metazoan macromolecular complexes. Nature 525: 339–344. https://doi.org/10.1038/nature14877

Wang Y., H. Zhang, B. Gigant, Y. Yu, Y. Wu, *et al.*, 2016 Structures of a diverse set of colchicine binding site inhibitors in complex with tubulin provide a rationale for drug discovery. The FEBS Journal 283: 102–111. https://doi.org/10.1111/febs.13555

Warner J. R., 1999 The economics of ribosome biosynthesis in yeast. Trends Biochem Sci 24: 437–440. https://doi.org/10.1016/s0968-0004(99)01460-7

Wawro M. E., K. Sobierajska, W. M. Ciszewski, W. Wagner, M. Frontczak, *et al.*, 2017 Tubulin beta 3 and 4 are involved in the generation of early fibrotic stages. Cellular Signalling 38: 26–38. https://doi.org/10.1016/j.cellsig.2017.06.014

Weems A., and M. McMurray, 2017 The step-wise pathway of septin hetero-octamer assembly in budding yeast, (Y. Barral, Ed.). eLife 6: e23689. https://doi.org/10.7554/eLife.23689

Weinberg D. E., P. Shah, S. W. Eichhorn, J. A. Hussmann, J. B. Plotkin, *et al.*, 2016 Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. Cell Rep 14: 1787–1799. https://doi.org/10.1016/j.celrep.2016.01.043

Weinstein B., and F. Solomon, 1990 Phenotypic consequences of tubulin overproduction in Saccharomyces cerevisiae: differences between alpha-tubulin and beta-tubulin. Molecular and Cellular Biology 10: 5295–5304. https://doi.org/10.1128/MCB.10.10.5295

Wickstead B., and K. Gull, 2007 Dyneins Across Eukaryotes: A Comparative Genomic Analysis. Traffic 8: 1708–1721. https://doi.org/10.1111/j.1600-0854.2007.00646.x

Wickstead B., K. Gull, and T. A. Richards, 2010 Patterns of kinesin evolution reveal a complex ancestral eukaryote with a multifunctional cytoskeleton. BMC Evolutionary Biology 10: 110. https://doi.org/10.1186/1471-2148-10-110

Wickstead B., and K. Gull, 2011 The evolution of the cytoskeleton. The Journal of Cell Biology 194: 513–525. https://doi.org/10.1083/jcb.201102065

Wickstead B., and K. Gull, Dyneins Across Eukaryotes: A Comparative Genomic Analysis. Traffic 8: 1708–1721. https://doi.org/10.1111/j.1600-0854.2007.00646.x

Winsor B., and E. Schiebel, 1997 Review: An Overview of the Saccharomyces cerevisiae Microtubule and Microfilament Cytoskeleton. Yeast 13: 399–434. https://doi.org/10.1002/(SICI)1097-0061(199704)13:5<399::AID-YEA126>3.0.CO;2-9

Winuthayanon W., M. L. Bernhardt, E. Padilla-Banks, P. H. Myers, M. L. Edin, *et al.*, Oviductal estrogen receptor α signaling prevents protease-mediated embryo death. eLife 4. https://doi.org/10.7554/eLife.10453

Winzeler E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, *et al.*, 1999 Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science 285: 901–906. https://doi.org/10.1126/science.285.5429.901

Wiśniewski J. R., A. Zougman, N. Nagaraj, and M. Mann, 2009 Universal sample preparation method for proteome analysis. Nature Methods 6: 359–362. https://doi.org/10.1038/nmeth.1322

Woods J. O., U. M. Singh-Blom, J. M. Laurent, K. L. McGary, and E. M. Marcotte, 2013 Prediction of gene-phenotype associations in humans, mice, and plants using phenologs. BMC Bioinformatics 14: 203. https://doi.org/10.1186/1471-2105-14-203

World Health Organization, 2018 *Preventive chemotherapy to control soil-transmitted helminth infections in at-risk... population groups.* WORLD HEALTH ORGANIZATION, S.l.

Xu D., Y. Pan, H. Zhang, X. Li, Y. Dai, *et al.*, 2015 Detection and characterization of carbendazim resistance in Sclerotinia sclerotiorum isolates from oilseed rape in Anhui Province of China. Genet. Mol. Res. 14: 16627–16638. https://doi.org/10.4238/2015.December.11.10

Xu X., D. Duan, and S.-J. Chen, 2017 CRISPR-Cas9 cleavage efficiency correlates strongly with target-sgRNA folding stability: from physical mechanism to off-target assessment. Sci Rep 7: 143. https://doi.org/10.1038/s41598-017-00180-1

Yamanouchi H., T. Umezu, and Y. Tomooka, 2010 Reconstruction of oviduct and demonstration of epithelial fate determination in mice. Biol Reprod 82: 528–533. https://doi.org/10.1095/biolreprod.109.078329

Yang P.-L., T.-H. Hsu, C.-W. Wang, and R.-H. Chen, 2016 Lipid droplets maintain lipid homeostasis during anaphase for efficient cell separation in budding yeast. Mol Biol Cell 27: 2368–2380. https://doi.org/10.1091/mbc.E16-02-0106

Yang F., S. Sun, G. Tan, M. Costanzo, D. E. Hill, *et al.*, 2017 Identifying pathogenicity of human variants via paralog-based yeast complementation. PLOS Genetics 13: e1006779. https://doi.org/10.1371/journal.pgen.1006779

Yang Y., M.-X. Li, Y.-B. Duan, T. Li, Y.-Y. Shi, *et al.*, 2018 A new point mutation in β2-tubulin confers resistance to carbendazim in Fusarium asiaticum. Pesticide Biochemistry and Physiology 145: 15–21. https://doi.org/10.1016/j.pestbp.2017.12.006

Yilmaz E., S. Ramünke, J. Demeler, and J. Krücken, 2017 Comparison of constitutive and thiabendazole-induced expression of five cytochrome P450 genes in fourth-stage larvae of Haemonchus contortus isolates with different drug susceptibility identifies one gene with high constitutive expression in a multi-resistant isolate. International Journal for Parasitology: Drugs and Drug Resistance 7: 362–369. https://doi.org/10.1016/j.ijpddr.2017.10.001

Yuan P., L. Zheng, H. Liang, Y. Li, H. Zhao, *et al.*, 2018 A novel mutation in the TUBB8 gene is associated with complete cleavage failure in fertilized eggs. J Assist Reprod Genet 35: 1349–1356. https://doi.org/10.1007/s10815-018-1188-3

Zeisel A., A. Yitzhaky, N. Bossel Ben-Moshe, and E. Domany, 2013 An accessible database for mouse and human whole transcriptome qPCR primers. Bioinformatics 29: 1355–1356. https://doi.org/10.1093/bioinformatics/btt145

Zhang Y.-J., J.-J. Yu, Y.-N. Zhang, X. Zhang, C.-J. Cheng, *et al.*, 2009 Effect of Carbendazim Resistance on Trichothecene Production and Aggressiveness of Fusarium graminearum. MPMI 22: 1143–1150. https://doi.org/10.1094/MPMI-22-9-1143

Zhang H., B. Brankovics, T. A. J. van der Lee, C. Waalwijk, A. A. D. van Diepeningen, *et al.*, 2016a A single-nucleotide-polymorphism-based genotyping assay for simultaneous detection of different carbendazim-resistant genotypes in the Fusarium graminearum species complex. PeerJ 4: e2609. https://doi.org/10.7717/peerj.2609

Zhang Z., R. B. Gasser, X. Yang, F. Yin, G. Zhao, *et al.*, 2016b Two benzimidazole resistance-associated SNPs in the isotype-1 β-tubulin gene predominate in Haemonchus contortus populations from eight regions in China. International Journal for Parasitology: Drugs and Drug Resistance 6: 199–206. https://doi.org/10.1016/j.ijpddr.2016.10.001

Zhao G., Y. Chen, L. Carey, and B. Futcher, 2016 Cyclin-Dependent Kinase Co-Ordinates Carbohydrate Metabolism and Cell Cycle in S. cerevisiae. Mol Cell 62: 546–557. https://doi.org/10.1016/j.molcel.2016.04.026

Zhu Z.-Q., F. Zhou, J.-L. Li, F.-X. Zhu, and H.-J. Ma, 2016 Carbendazim resistance in field isolates of Sclerotinia sclerotiorum in China and its management. Crop Protection 81: 115–121. https://doi.org/10.1016/j.cropro.2015.12.011

Zotenko E., J. Mestre, D. P. O'Leary, and T. M. Przytycka, 2008 Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. PLoS Comput Biol 4: e1000140. https://doi.org/10.1371/journal.pcbi.1000140

Vita

Riddhiman Kannan Garge grew up in Bangalore, India. He completed his undergraduate degree from M. S. Ramaiah Insititute of Technology while parallelly pursuing research at the National Center for Biological Sciences (NCBS). Subsequently, he joined the Biochemistry graduate program at University of Texas at Austin. His research areas include structural biology at NCBS, *in silico* drug design against cancer targets at Schrödinger Inc., and evolution, systems, and synthetic biology at the University of Texas at Austin.

Permanent email: riddhiman92@gmail.com

This dissertation was typed by the author.